

UNIVERSIDADE FEDERAL DO PARANÁ
DEPARTAMENTO DE CIÊNCIA E GESTÃO DA INFORMAÇÃO

IDEMARA MARCELI MARTINS

**A MINERAÇÃO DE DADOS PARA DESCOBERTA DE CONHECIMENTO E
UMA OFERTA ADEQUADA NO CANAL DE TELEVISÃO ABERTA**

CURITIBA
2010

IDEMARA MARCELI MARTINS

**A MINERAÇÃO DE DADOS PARA DESCOBERTA DE CONHECIMENTO E
UMA OFERTA ADEQUADA NO CANAL DE TELEVISÃO ABERTA**

Monografia apresentada como requisito a obtenção do diploma de Bacharel em Gestão da Informação ao curso de graduação em Gestão da Informação do Setor de Ciências Sociais Aplicadas da Universidade Federal do Paraná.

Orientadora: Prof^a Dra Denise Fukumi Tsunoda

CURITIBA
2010

RESUMO

A pesquisa viabiliza a técnica de Mineração de Dados como ferramenta alternativa de sistema de informação transformando dados em informação para tomada de decisão em negócios no processo de Descoberta de Conhecimento em Bases de Dados possibilitando o relacionamento de metodologias de *Business Intelligence* e Inteligência Competitiva. A aplicação gerou uma amostra intencional, de cunho descritivo e quantitativo realizado no Setor de Ciências Sociais Aplicadas da Universidade Federal do Paraná, descrevendo o perfil do universitário como telespectadores para TV aberta local. Utiliza uma técnica, o algoritmo *Apriori*, para descobrir regras de associação nos conjuntos de dados coletados com o objetivo de descrever dados, identificar padrões e possibilitar oferta para um público específico ou marketing direcionado. Com aplicação de estudo de caso na Band Curitiba, o trabalho identificou as necessidades informacionais da empresa referente a clientes pela observação e referencial teórico. E como produto final apresenta o modelo de conhecimento para a empresa integrada na pesquisa como contribuição a tomada de decisão com técnica de inteligência nos negócios.

Palavras-chave: *Business Intelligence*; Gestão da Informação; Informação para Negócios; Inteligência Artificial; Inteligência Competitiva; TV aberta

LISTA DE SIGLAS

AM	Aprendizagem de Máquina
IA	Inteligência Artificial
BI	Business Intelligence
MD	<i>Data Mining</i>
GC	Gestão do Conhecimento
IC	Inteligência Competitiva
KDD	<i>Knowledge Discovery in Databases</i>
RNA	Rede Neural Artificial
WEKA	<i>Waikato Environment Knowledge Analysis</i>

SUMÁRIO

1 INTRODUÇÃO	5
1.1 DESCRIÇÃO DO PROBLEMA	7
1.2 OBJETIVOS	8
1.2.1 Objetivo Geral	8
1.2.2 Objetivos Específicos	8
1.3 JUSTIFICATIVA	9
2 DESCOBERTA DE CONHECIMENTO EM BASES DE DADOS: REFERENCIAL TEÓRICO	11
2.1 A CONFLUÊNCIA DO DATA MINING	13
2.1.1 Inteligência Artificial	15
2.1.2 Business Intelligence	17
2.1.3 Gestão do Conhecimento	19
2.2 MINERAÇÃO DE DADOS	22
2.2.1 Tarefas de Mineração de Dados	22
3 FASES PROCESSO KDD	28
3.1 IDENTIFICANDO OS DADOS	28
3.2 COLETA DE DADOS	29
3.3 PRÉ PROCESSAMENTO	30
3.4 TÉCNICA DE ASSOCIAÇÃO UTILIZADA	31
3.4.1 O Algoritmo Apriori	32
4 METODOLOGIA	35
4.1 DELINEAMENTO E DELIMITAÇÃO DA PESQUISA	35
4.2 AMBIENTE DO ESTUDO DE CASO	35
4.3 APRESENTAÇÃO DAS VARIÁVEIS	37
4.4 POPULAÇÃO E AMOSTRAGEM	38
4.5 COLETA DE DADOS	39
4.6 SISTEMATIZAÇÃO E ANÁLISE DOS DADOS	41
4.7 FERRAMENTA WEKA	46
5 RESULTADOS: ANÁLISE E DISCUSSÃO	48
5.1 RESULTADOS OBTIDOS PELOS QUESTIONÁRIOS	48
5.2 RESULTADOS OBTIDOS PELA MINERAÇÃO DE DADOS – FERRAMENTA WEKA	56
5.3 ANÁLISE DAS REGRAS GERADAS	57
6 CONSIDERAÇÕES FINAIS	64
6.1 ALGUMAS CONTRIBUIÇÕES	66
REFERÊNCIAS	67
APÊNDICE A FORMULÁRIO – PERFIS DE TELESPECTADORES DE TV ABERTA LOCAL ENTRE UNIVERSITÁRIOS DO SETOR DE CIÊNCIAS SOCIAIS APLICADAS DA UFPR	70

1 INTRODUÇÃO

Os métodos automáticos para proporcionar análises em bases de dados surgem da necessidade de extrair conhecimentos contidos em grandes volumes de dados que propiciam a descoberta de conhecimento por meio de uma aplicação computacional facilitadora e precisa.

Dados sobre clientes são um recurso necessário para empresas que buscam adequar seus produtos e serviços relativos aos interesses exigidos ou previamente caracterizados. Recursos tecnológicos possibilitam uma vasta capacidade de armazenamento e métodos de aplicação para análise, ou seja, um campo de estudo conhecido como *Knowledge Discovery in Databases* (KDD) – Descoberta de Conhecimento em Bases de Dados (FAYYAD, 1996). Um processo aplicado em bases de dados, métodos estatísticos, ferramentas de visualização e técnicas de Inteligência Artificial (IA), dividindo-se em fases de seleção, pré-processamento, mineração de dados e avaliação/interpretação. Sendo, a Mineração de Dados a fase mais importante de um trabalho que busca métodos de automação especializados para aprendizado a partir de dados.

A Mineração de Dados (DM), ou *Data Mining*, é um processo que descobre informações relevantes, como padrões, associações, mudanças, anomalias e estruturas em dados armazenados em banco de dados. Tal processo pode surgir de algumas dificuldades ou problemas como pouca qualidade em dados mal estruturados, falta de condição sobre dados, ausência ou perda de informação armazenada ou mesmo para exploração de conjuntos de dados com um propósito específico de análise. Para minimização dessas perdas, tem-se a fase do pré-processamento na qual os dados são preparados buscando melhorar a qualidade das informações armazenadas.

O pré-processamento é uma fase importante que precede o *Data Mining*, pois trata os valores contidos nas bases e também da ausências destes. Se o domínio a ser estudado for a renda mensal de um cliente objetivando a identificação de sua classe social, nem todos os cadastros apresentam essa informação por precaução do próprio cliente em fornecer tal dado, por exemplo. Sem o pré-processamento os resultados podem ser seriamente afetados, imprecisos ou inválidos. Outra justificativa é quando a maioria dos atributos não é independente entre si. Ou seja,

uma correlação a ser devidamente tratada exigindo uma substituição por valores estimados denotando o atributo permitindo sua independência e selecionando um método mais apropriado para tal conjunto de dados (BATISTA, 2003).

A pesquisa sobre perfis de telespectadores universitários para TV aberta local visa aplicar métodos de Descoberta de Conhecimento com tarefas de associação para tratar de dados sobre possíveis telespectadores ou público definido para a TV Band Curitiba canal 2. Esta é uma forma de apresentar ocorrências reais e observadas para descrição de dados a fim de apoiar futuras estratégias de programação, retenção de clientes e até possíveis conquistas de novos nichos de mercado ou serviços.

Para a análise de Mineração de Dados e a apresentação de seu conhecimento descoberto o especialista deve corresponder com as exigências do KDD obedecendo todas suas fases: o pré-processamento após a coleta de dados e sua organização; mineração de dados; e pós-processamento, ou seja, o nível satisfatório do tratamento da informação para resultados. Para aplicação da técnica, foi utilizado o método de associação. Trata-se de uma tarefa clássica associa itens de ocorrência freqüente em um conjunto de dados.

O trabalho também é composto por mais seis seções. Em sua segunda seção “Descoberta de Conhecimento em Bases de Dados: referencial teórico” aborda desde o surgimento do processo KDD com necessidade de utilização, a confluência da Mineração de Dados com algumas disciplinas e suas tarefas. A terceira seção descreve as fases do processo de Descoberta de Conhecimento em Bases de Dados e a técnica de Mineração utilizada para tratamento dos dados nesta pesquisa. Na quarta seção, a metodologia de pesquisa utilizada para realização deste trabalho. A quinta seção descreve o estudo de caso da TV Band Curitiba. Na sexta seção são apresentados os resultados da pesquisa bem como a ferramenta WEKA para tratar os valores contidos na base construída para esta finalidade de observação e a utilização da tarefa associativa de Data Mining na técnica do Apriori. O sétimo capítulo traz conclusões incluindo dificuldades encontradas e a finalização das referências.

1.1 DESCRIÇÃO DO PROBLEMA

A análise em perfis de telespectadores é para que a TV Band Curitiba busque apresentar programas e serviços adequado às preferências de um público específico, o que pode justificar que seu nível de audiência não está na posição de liderança em alguns horários nas divulgações nacionais que baseiam as regiões ligadas a *share* (UOL NOTÍCIAS). A ausência um conjunto de dados referente a clientes pode dificultar o conhecimento de preferências ou receptividade de seu produto/serviço.

A TV Bandeirantes do Paraná – TV Band busca oferecer ao cliente e telespectador proposta de apresentação pela sua credibilidade regional, qualidade profissional e apoio justificado e aprovado em porcentual de receptividade. Porém, sem o conhecimento criterioso atualizado de quem a assiste, a procura por seu serviço bem como um relacionamento e/ou acesso direto ao cliente telespectador será impedido ou substituído pelo serviço concorrente. Surge, assim, uma oportunidade tanto para a TV como para a comunidade em participar de uma possível personalização do produto no tocante aos seus interesses e atendimento social. Para que isto ocorra é essencial a produção adequada ou a reconstrução do conteúdo audiovisual como um efetivo canal de televisão para o serviço público tanto para uma adequação ou produção participativa de sua programação como para a realização dos direitos fundamentais adequado ao regime privado, ou seja, para o mercado.

Um veículo de radiodifusão de imagens que transmite programas de entretenimento, educação e informação, obtém-se de uma recepção simultânea pelos telespectadores. Se tal procedimento não ocorrer de forma eficiente e condizente com exigências dos clientes acaba gerando prejuízos e até a perda destes.

Entretanto, com justificativa de marketing observada, pode-se afirmar que houve uma decadência na receptividade da TV Band em sua transmissão em tempos remotos gerando, com isso, um impacto significativo para vendas de sua cobertura e credibilidade tradicional.

As organizações precisam agir com inteligência, monitorar informações ambientais e principalmente seu público alvo. Torna-se imprescindível o investimento em tecnologia em um ambiente em que tudo é correspondido por

bases eletrônicas propensos a respostas com valor numérico preciso baseado em ocorrências reais. A falta de sistematização de dados referentes ao negócio torna a organização impedida de responder competitivamente aos desafios e oportunidades que se apresentam no setor.

Diante do contexto da organização algumas questões precisam ser resolvidas neste trabalho: “O que fazer com armazenamento de dados relativos a perfis de clientes?”, “Como os dados digitais podem trazer benefício para a empresa?”, “Como tornar útil um conjunto de dados de dados?”

Essas perguntas remetem a uma única questão de pesquisa, ou seja: **Como associar informações de clientes contidos em conjuntos de dados de forma a analisar e utilizar de maneira útil a informação para negócios por meio da prática de Mineração de Dados no processo de Descoberta de Conhecimento em Bases de Dados de um canal de televisão aberta?**

1.2 OBJETIVOS

Neste item são apresentados o objetivo geral e os objetivos específicos da pesquisa.

1.2.1 Objetivo Geral

O objetivo geral do trabalho é comportar clientes telespectadores universitários com perfis apresentados em dados estruturados para descoberta de padrões em associações que expressem informação útil à tomada de decisão e geração de conhecimento com auxílio da aplicação de técnica de Mineração de Dados.

1.2.2 Objetivos Específicos

- identificar os tipos de telespectadores universitários e descrever seus perfis de acordo com suas preferências para TV;

- analisar se a programação de TV é adequada às preferências e disponibilidade dos telespectadores;
- Verificar o acesso de telespectadores a canais de TV diferenciados e/ou específicos;
- analisar o nível de frequência diária dos telespectadores na TV utilizando-se de atributos em regras para Mineração de Dados para obter informação de valor futuro ou desconhecido.

1.3 JUSTIFICATIVA

A finalidade principal da análise em perfis de telespectador é para que a TV Band Curitiba apresente programas e serviços adequados à preferência do seu público alvo. Também, o interesse da pesquisadora em apresentar resultados satisfatórios à empresa, sendo vinculada a esta como funcionária comprometida com a missão da organização em atingir seus objetivos e/ou contribuir para o sucesso do seu negócio.

Levando-se em conta que a Mineração de Dados tem por objetivo apresentar regras associativas com base em ocorrências freqüentes, esta oferece conclusões tanto para riscos a enfrentar como para novas oportunidades e inovação. A identificação de atributos em perfis levantados e o tratamento dos dados referente a essa questão, busca a informação para fornecer subsídios para novas ações para o setor de comunicação a fim de que a empresa atue competitivamente.

Desde o ano de 2009, a TV Bandeirantes do Paraná (Band Curitiba), vem modernizando seu quadro na segmentação regional. Considerando seu porte empresarial, a Band Curitiba constitui-se de eficientes ferramentas e/ou estrutura propícios ao controle do negócio com foco em marketing e/ou base de dados para avaliação de seu público para tomada de decisão nos negócios. Tal subsídio torna a empresa capaz de atuar sobre as distorções possíveis de serem descobertas em diferentes perfis a partir de dados dispostos neste sentido, para informação e conhecimento.

A BI (*Business Intelligence* - Inteligência nos Negócios) é infra-estrutura que alinha TI com práticas de negócio, tem a capacidade de estruturar histórico de dados

como forma alternativa de para tratamento de informações relevantes para o negócio empresarial. Informação para negócios é estimulada pela competitividade (atividades dos concorrentes e do mercado em geral), esta inclui técnicas de observação e acompanhamento constantes de dados e conhecimento relevantes ao objetivo da organização. A Inteligência Competitiva adéqua a empresa para tal modelo de concorrência empurrando-a para a inovação com apoio de recurso tecnológico para gerar novas idéias, resolver problemas e/ou auxiliar a tomada de decisão de maneira a tornar esta a frente ou em paralelo a concorrência.

Para comportamento de uma base de dados favorecida à empresa de estudo de caso, foi escolhida a população-alvo no Setor de Ciências Sociais Aplicadas da UFPR. Esta amostra deve-se à proximidade da pesquisadora no campus e pela possibilidade de universitários representarem uma classe social e um público definido para um segmento de mercado com potencial consumo e/ou utilização dos serviços de TV.

Com esta pesquisa espera-se garantir que a TV Band Curitiba tenha mais conhecimento de um público específico e que possa contribuir para o desenvolvimento ou aprimoramento de seu sistema de informação para negócios com gestão em Inteligência Competitiva com dados propícios a aplicação de técnica de Mineração de Dados.

2 DESCOBERTA DE CONHECIMENTO EM BASES DE DADOS: REFERENCIAL TEÓRICO

A descoberta de conhecimento é uma atividade de gestão que pode ser considerada um processo que envolve planejamento, organização, distribuição e controle como um recurso sobre dados, por exemplo, visando à racionalização e efetividade de um sistema, produto ou serviço (MANUAL..., 1997). Neste sentido, a informação percorre alguns níveis desde sua descoberta ou estado primário até sua representação para utilização como conhecimento, o que pode ser notado na tradicional pirâmide da informação que assume três propriedades de abstração em seu processo, como apresenta a Figura 1:



FIGURA 1. HIERARQUIA ENTRE DADO, INFORMAÇÃO E CONHECIMENTO
FONTE: (GOLSCHMIDT e PASSOS, 2006).

Davenport (1998) apresenta os níveis da pirâmide isolados entre si, mas em sua estrutura é integrada à informação em estado de equilíbrio como um processo. A base da pirâmide representa simples observações ou itens que podem ser captados, observados por pessoas ou armazenados por recursos da Tecnologia da Informação facilmente estruturados, freqüentemente quantificados e transferíveis. No nível intermediário, está a informação que representa os dados processados em unidade de análise, com significado e exigência de consenso e mediação humana. Por fim, o conhecimento é a informação com análise no estágio de reflexão, síntese e contexto. Choo (2002) observa essas etapas em perspectiva de um *continuum* de valores fundamentalmente identificados pela contribuição humana para resultados,

capacitação organizacional e para criação e compartilhamento de conhecimento como em repetidos ciclos de processamento de informação pela gestão, ação, resultado e retroalimentação desta a fim de criar valor para seus clientes e a sociedade, explorar oportunidades, evitar ou enfrentar crises e ameaças, além de tornar organizações competitivas.

Uma das preocupações de uma organização é enfrentar os desafios concorrenciais e manter sustentabilidade e fidelidade dos seus clientes pelas suas preferências. Para isso, as empresas devem aumentar o conhecimento do seu negócio melhorando seus processos, competitividade e resultados, explorar informação disponível ou buscar dados que formule e proporcione respostas a questões de negócio mais complexas. Observa-se que a Mineração de Dados possibilita a descoberta de padrões freqüentes e apresenta a informação útil para auxílio à tomada de decisão de empresas, aperfeiçoa métodos de produção e encontra novas oportunidades de negócios.

Trata-se de uma área denominada Descoberta de Conhecimento em Bases de Dados (*Knowledge Discovery in Databases – KDD*). A denominação do processo de descoberta de conhecimento em bases de dados é chamada popularmente de Mineração de Dados, embora MD seja apenas a técnica aplicada. Segundo Fayyad et al. (1996^a), “KDD é um processo, de várias etapas, não trivial, interativo e iterativo, para identificação de padrões compreensíveis, válidos, novos e potencialmente úteis a partir de grandes conjuntos de dados”.

A Mineração de Dados é uma técnica que visa buscar conhecimento a partir dos dados que envolve e relaciona a informação formando um conjunto de padrões apresentados na forma de regra: SE <condições> ENTÃO <conclusões>, por exemplo (GOLDSCHMIDT e PASSOS, 2005). Ou seja, surge a possibilidade de otimizar a informação como recurso associado a um valor, que implica um envolvimento consistente, selecionado e avaliado no gerenciamento e utilização relacionados ao trinômio dado, informação e conhecimento para a Gestão da Informação (MARCHIORI, 2002).

Para GOLDSCHMIDT e PASSOS (2005), a aplicação de KDD é utilizada tanto para orientação das ações a serem realizadas quanto ao macro objetivo pretendido. Na primeira o especialista da área se dispõe de hipóteses que devem ser refutadas ou comprovadas mediante a análise de dados e busca efetiva de conhecimento a partir da abstração dos dados existentes. Na segunda, com

pretensão macro objetiva, permite prever valores de determinados atributos em novas situações a partir do histórico de casos verificados em conjunto de dados e descreve, de forma compreensível para homem, o conhecimento intrínseco nessa base.

Entre diversas aplicações possíveis para a Descoberta de Conhecimento em Bases de Dados, casos relacionados com clientes com foco em predição de lealdade, descrição de perfis, análise de comportamento e retenção é uma das aplicações mais importantes ou de sucesso do KDD (KLÖSGEN et al., 2002).

Embora sempre houvesse a necessidade dessas informações estarem presentes para o acesso empresarial, tal importância torna-se preeminente com adventos da competitividade. Na tomada de decisões empresariais, a informação para negócios é usada para redução de incertezas, monitoração da concorrência, identificação de ameaças e oportunidades e melhoria da competitividade. A competitividade entre empresas pode estar vinculada à qualidade do sistema de informação de que se dispõe em relação a concorrentes (BARRETO, 1996).

No entanto, o diferencial competitivo é justamente a aplicação da técnica certa para cada tipo de demanda, bem como o profissional envolvido no processo de Descoberta de Conhecimento em Bases de Dados. Transformação da informação em conhecimento de forma inteligente representa um importante fator crítico de sucesso na empresa, a fim de que o emprego da tecnologia adequada seja um correto investimento em termos de tempo, recursos materiais e financeiros, (DAVENPORT e PRUSAK, 1998).

2.1 A CONFLUÊNCIA DO DATA MINING

A Mineração de Dados surge a partir de todas as fases de processamento constituída sobre metodologia e algoritmos que os pesquisadores desenvolveram para aprimorar técnicas mais eficientes para busca, tratamento e recuperação da informação (TAN et.al., 2006). Assim, integra idéias como: amostragem, estimativa e teste de hipóteses a partir de estatísticas; algoritmos de busca, técnicas de modelagem e teorias de aprendizagem da Inteligência Artificial, reconhecimento de padrões e aprendizagem de máquina; além de adotar idéias de outras áreas, incluindo otimização, computação evolutiva, teoria da informação, processamento de

sinais, visualização e recuperação de informações. Tal relacionamento da Mineração de Dados com outras áreas pode ser representado na Figura 2:

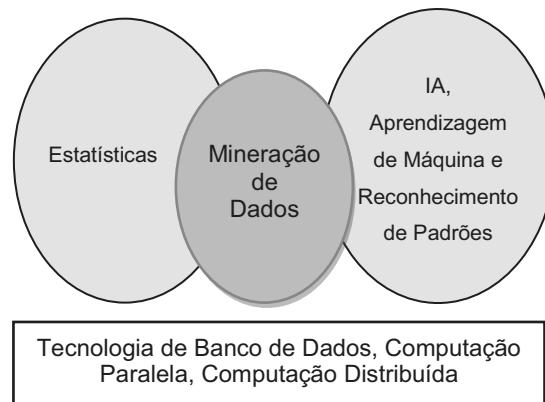


FIGURA 2 MINERAÇÃO DE DADOS COMO UMA CONFLUÊNCIA DE MUITAS DISCIPLINAS.
 FONTE: (TAN et. al 2006, p.8)

Atualmente, o mercado de mineração de dados tem crescido consideravelmente, mas existem poucas ferramentas de KDD bem aplicadas. Assim, surge o interesse de desenvolver bases com dados reais, propensos a sistemas de produção e com históricos ou variáveis para qualificação (FAYYAD et al., 1996a).

Outro exemplo para confluência do *Data Mining* pode ser o *Knowledge Management System* (Sistema de Gestão do Conhecimento). Barbieri (2001, p.5), diz que é uma fase de transformação do gerenciamento do conhecimento, que toma a informação como suporte para tomada de decisão menos compartimentada e estruturada como é o Business Intelligence (BI – Inteligência nos Negócios). A área do Business Intelligence oferece um conjunto de metodologias e ferramentas voltadas para a obtenção, armazenamento e análise de dados para obtenção de informação relevante para descoberta de conhecimento e indicadores de desempenho.

A organização de áreas do conhecimento estabelece uma ponte entre diferentes abordagens para tratamento da informação, fazendo com que a Mineração de Dados seja uma ferramenta com bases para pesquisa em diferentes campos de estudo e difundidas como a Inteligência Artificial, *Business Intelligence* e a Gestão do Conhecimento. O papel do *Business Intelligence* seria disponibilizar as informações, enquanto a Gestão do Conhecimento seria as combinações e

disseminações de forma a transformarem a informação em conhecimento, ou seja, uma forma de administração, compartilhamento e aproveitamento do conhecimento das pessoas na disseminação das melhores práticas para o crescimento da organização (REZENDE, 2003). As tecnologias poderão se fundir acoplando e disponibilizando ferramentas inspirando-se para outros conceitos.

[...] A gestão do conhecimento ou Knowledge Management (KM) e a inteligência de negócios ou Business Intelligence (BI), aplicados juntamente com os recursos dos Sistemas de Informação e da Tecnologia da Informação, podem facilitar a geração de informações oportunas para favorecer os decisores empresariais em suas atividades estratégicas. [...] (REZENDE, ABREU. 2003, p.221)

As abordagens se confundem ou se apropriam de definições de diversas áreas de estudo. O tratamento da informação tem sido considerado um bem ativo dentro das empresas exigindo, assim uma utilização inteligente para negócios desenvolvendo ferramentas inovadoras com Mineração de Dados para melhorar resultados criando similaridades e condições de relação para se adequar aos próprios modelos de negócios existentes. As técnicas de MD são tradicionalmente reconhecidas em técnicas de Inteligência Artificial para aprendizado de máquina, conforme conceituado posteriormente.

2.1.1 Inteligência Artificial

A Inteligência Artificial – IA, surgiu de tarefas simples do raciocínio humano em suas tarefas cotidianas para objetos físicos e relacionamentos para aplicações lógicas em manipulações simbólicas básicas. Para Rich e Knight (1993), seu desenvolvimento é aprimorado com manipulação de suas técnicas com a utilização de maior quantidade de conhecimento incluindo percepção (visão e fala), a compreensão da linguagem natural e a solução de problemas em domínios especializados, como diagnóstico médico e análise química. Neste sentido, o ser humano pode ser diferenciado por tais habilidades de comunicação e relacionamento como *inteligência*. Assim, a utilização de técnicas semelhantes ao natural em aplicações similares é definida como *Inteligência Artificial* na manipulação computacional, ou seja, fazer com que computadores realizem coisas que, no momento, as pessoas fazem melhor Rich e Knight (1993).

Ainda em desenvolvimento, a IA está se sustentando como disciplina prática pela sua aplicação e pelos resultados dos sistemas especializados em operações cotidianas nas organizações sem a necessidade de se justificar no conhecimento do senso comum, a pesquisa.

IA: Seus benefícios para Descoberta de Conhecimento em Bases de Dados

A atribuição dos símbolos para identificação de situações é comum tanto na IA como aplicações de Mineração de Dados. Rich e Knight (1993) afirmam que difíceis resultados de pesquisa em IA é que esta requer conhecimento, ou seja, explora conhecimento identificado com as seguintes propriedades:

- é volumoso;
- é difícil caracterizá-lo com precisão;
- muda constantemente;
- difere de simples dados por organizar-se de uma maneira que corresponde ao modo como será usado.

Os mesmos autores explicam que aplicação de aprendizagem de máquina tem sido voltada para abordagens analíticas porque é rica em dados e em conhecimento. O sistema aprende a partir de um exemplo fornecido, e o resultado é disponibilizado por causa desse conhecimento. Tal exploração captura generalizações em situações que compartilham propriedades importantes e sejam agrupadas; por fim, possam ser modificadas para correção de erros e reflexão de mudanças no contexto referido; e até auxiliando na possibilidade de limitar ou superar o próprio volume de dados.

O Aprendizado de Máquina tem seu conceito dado à seguinte função de três fatores: a tolerância de erros (h), o número de características binárias presentes nos exemplos (t), e o tamanho da regra necessária para fazer a discriminação (f). Se os números de exemplos de treinamento forem exigidos for polinomial em h , t e f , então diz-se que o conceito é aprendível (Rick e Knight, 1993). A teoria é uma técnica igualmente aplicada em Mineração de Dados pelo conceito dado à teoria da aprendizagem.

2.1.2 Business Intelligence

A Inteligência nos negócios (BI) possibilita que as organizações realizem uma série de análises e projeções, agilizando seus processos sempre relacionados às tomadas de decisão, como um diferencial para a empresa, ou seja, uma ferramenta diferenciadora para mercado. Não é obrigatória a existência da Mineração de Dados no BI, é uma proposta de abordagem devido ao seu conceito e tratamento dos dados e informações.

Histórico e conceito geral de Business Intelligence

A busca por novas metodologias ocorreu na década de 80 com a chegada dos bancos de dados relacionais e interfaces gráficas voltadas para urgentes necessidades empresariais para negócios. O conceito de *Business Intelligence* é mais antigo ainda. Segundo Ribeiro et al (2003, p.1-2):

[...] A história do BI que conhecemos hoje, começa na década de 70, quando alguns produtos de BI foram disponibilizados para os analistas de negócio. O grande problema era que esses produtos exigiam intensa e exaustiva programação, não disponibilizavam informação em tempo hábil nem de forma flexível, e além de tudo tinham alto custo de implantação [...].

Segundo Barbieri (2001), na década de 90 tornou-se inevitável a freqüente armazenagem de dados somando-se ao advento da internet gerando novos métodos de tratamento e comunicação diretamente ligados aos negócios da empresa. A chamada Revolução Tecnológica:

[...] Diferentemente da Revolução Industrial, onde as benesses não chegavam a todos diretamente, a revolução dos dados e da informação, capitaneada pela Internet, tornou-se democrática, invasiva e de amplo alcance.[...] (BARBIERI. 2006, p.3)

A partir da influencia de métodos computacionais surge a necessidade das empresas manterem seu espaço no mercado e a fidelidade de seus clientes com processos estratégicos e inteligentes da informação necessária, ou seja, gerou uma busca por ferramentas que disponibilizem informações de forma segura, completa e que possibilite uma interação com os dados. Surgindo o conceito de Mineração de

Dados e a possibilidade de simulações com dados históricos. Conforme Barbieri (2006, p.5), “chegamos a era do BI - Business Intelligence, com a informação usinada a partir dos dados sendo diretamente aplicada aos negócios[...]”. O *Business Intelligence* está dentro do conceito de Gestão do Conhecimento que será apresentado posteriormente.

Os benefícios da ferramenta de Business Intelligence

Os dados podem estar dispostos em repositórios simples com dados históricos e outros dados internos e externos úteis aos negócios da empresa. Como a Mineração de Dados apresenta resultados variáveis, dentro do BI torna-se uma ferramenta favorável quando se pretende assimilar perfis de seus clientes referentes à suas preferências que apontam com a oferta de um concorrente, por exemplo. Segundo Mylius (2003, p.54) diz que “as necessidades de geração de informações mais conhecidas são aquelas referentes a clientes, faturamento, margem e custos [...]”.

Como já citado, a oferta de acordo com o perfil do cliente é um dos fatores necessários à sobrevivência da empresa. O *Business Intelligence* também pode apresentar informações para diversas áreas passíveis de sua aplicabilidade com análises matemáticas, financeiras e estatísticas, seu principal objetivo é analisar a informação e decidir a partir dela. Tal tipo de abordagem é um diferencial competitivo em meio a um cenário global de concorrência em que as empresas estão atualmente, ou seja, estão sendo forçadas a uma abordagem pró-ativa para bons programas de marketing. A MD possibilita que seja visualizado um cruzamento de dados de insatisfação ou recusa por produtos a partir de relatórios adequados para possibilitar relacionamento com clientes, por exemplo. Ou ainda, de forma previsível, verificar quais clientes trocaria pelo serviço concorrente.

[...] O que importa, no fundo, é o cultivo de um bom relacionamento com o elemento principal do circuito, que é o cliente. O que se vende, e como se vende parecem ganhar status de menor importância do que o fato de poder estar mais próximo dele: sua majestade o cliente. (BARBIERI. 2001, p.11)

O negócio de uma empresa é a relação do cliente com o fornecedor. E a Inteligência nos Negócios é a decisão correta da empresa com o objetivo de atender

satisfatoriamente o cliente. Isto é como o conhecimento empírico que tira o melhor proveito das informações dispostas para serem empregadas nos negócios.

2.1.3 Gestão do Conhecimento

A definição da Gestão do Conhecimento está baseada em duas importantes definições de conhecimentos para uma organização. Segundo Nonaka (1997) o conhecimento é formal ou explícito, quando contido nos livros, manuais, documentos, periódicos, base de dados, repositórios, entre outros; e o informal ou tácito, que é gerado e utilizado no processo de produção do conhecimento formal, constituído de idéias, fatos, suposições, decisões, questões, experiências e pontos de vistas, sendo este considerado como um ativo de patrimônio de grande valor.

Segundo Terra (2003) Gestão do Conhecimento (GC) é a capacidade que as organizações têm em criar e utilizar o conhecimento através de seus processos. De acordo com Rezende (2003) informações relevantes e úteis “geradas e trabalhadas” por pessoas e/ou recursos computacionais, produzidas com qualidade e de forma antecipada, podem ser chamadas de conhecimento. A Gestão do Conhecimento pode ser entendida como uma forma de administração, compartilhamento e aproveitamento do conhecimento das pessoas na disseminação das melhores práticas para o crescimento da organização (REZENDE, 2003). Ou seja, o trabalho deve ser em conjunto a fim de facilitar o aprendizado coletivo planejando pessoas, cultura, processos e tecnologia para desenvolver e estruturar a organização para a geração do conhecimento. GC controla todo tipo de conhecimento contido numa organização a partir de seu conjunto de atividades (estabelecimento de políticas, procedimentos, tecnologias) capazes de coletar, distribuir e utilizar efetivamente o conhecimento, (MORESI, 2001). Assim, a Gestão do Conhecimento considera a Mineração de Dados como uma ferramenta de auxílio que descobre conhecimento contido em grande conjunto de informações em dados. Isso resolve o desafio da GC em transformar as bases de informações da organização em conhecimento para negócios. As práticas da Gestão do Conhecimento são a Inteligência Competitiva, Mapeamento de Competências, Aprendizagem Organizacional, *Benchmarking*, *Coaching*, Comunicação Institucional, Comunidades de Prática, Educação Corporativa/Universidade Corporativa, Gestão de Marcas e Patentes, Gestão do

Relacionamento com os Clientes, Lições Aprendidas Mapeamento de Competências, Mapeamento de Processos, Melhores Práticas, Memória Organizacional, *Mentoring*, Normalização e Padronização e Portais Corporativos.

Inteligência Competitiva: seus benefícios com *Data Mining*

A abordagem de estudo de *Data Mining* com Inteligência Competitiva (IC) chama uma empresa de *organização inteligente* por tratar a informação nela existente para negócios com aplicação de mecanismos que utilizem os dados e apresentem uma solução para decisão interna em resposta ao ambiente externo que a IC visa monitorar. Esta se torna uma dificuldade para a empresa que busca aumentar seu nível de utilização da informação com subsídios insuficientes para tratamento de grandes quantidades de dados e/ou dados com padronização complexa de assimilar convencionalmente.

A MD não necessariamente está vinculada a IC para Gestão do Conhecimento ou Gestão da Informação para que as organizações sejam inteligentes sendo estas duas ferramentas independentes em cada campo de conhecimento e sua junção esclarece e melhor adéqua a situação da organização na aplicação científica. A coleta de dados é uma solução possível para obtenção de informação do ambiente externo ou cliente em potencial a fim de que isto seja monitorado, armazenado e processado para tomada de decisão e vantagem competitiva (TERRA & ALMEIDA, 2008). Isso se chama monitoramento ambiental para a IC, o que pode ser percebido no processo de implementação da Mineração de Dados que apresenta a informação sobre dados como forma de agregar valor aos dados ou aprender com a informação obtida (ALMEIDA & COSTA, 2008). Para Terra e Almeida (2008), é necessário algumas reflexões sobre os processos de implementação da IC na empresa, composta de algumas fases, compondo um fluxo desse processo que envolve planejamento e identificação das necessidades da informação, coleta e tratamento da informação, análise e validade da informação, disseminação e utilização estratégica da informação e avaliação pelos responsáveis pela unidade de inteligência.

A intenção de coleta e seleção da informação é de que sua utilização seja uma ferramenta apropriada para geração de conhecimento disseminada para os

colaboradores responsáveis pelo negócio da empresa e inovação do seu produto (TERRA, 2003). Isto está fortemente relacionado a utilização de recursos de Tecnologia da Informação, o que será apresentado a seguir.

Tecnologia da Informação e Inteligência Competitiva

O processamento de dados com a complexidade da utilização da informação obriga que a computação especializada atenda as necessidades empresariais de adquirir conhecimento ocultos em bases de dados de difícil manipulação convencional (REZENDE, 2006). Conforme Terra (2003), a empresa pode buscar tecnologia que sustente o processamento e analise a informação. A Tecnologia da Informação (TI) pode apoiar os processos de IC, organizando e tratando a informação para auxílio nos principais objetivos do sistema de Inteligência Competitiva. Isto se apresenta na Figura 3:



FIGURA 3 GESTÃO DO CONHECIMENTO E INTELIGÊNCIA COMPETITIVA.
FONTE: Biblioteca TerraForum Consultores (2008, p.2)

Ferramentas de TI destinadas ao apoio da IC

Surgem ferramentas de TI que apoiam a aplicação da IC atendendo a informação em seus processos de: busca, organização e disseminação da

informação. Neste caso é o ambiente que torna favorável o acesso e armazenamento organizado do conteúdo estratégico com a utilização da ferramenta utilizada neste trabalho (CARVALHO, 2005):

- *Business Intelligence*: área que oferece um conjunto de metodologias e ferramentas que possibilitam a análise de dados para obtenção de informação em dados obtidos e armazenados como indicadores de desempenho empresarial. Entre as tecnologias pode citar:

- *Data Mining*: técnica que descobre informações relevantes em dados armazenados em bancos de dados.

Inteligência empresarial é um processo sistêmico de coleta, análise e distribuição de dados para melhorar a decisão dos negócios. Nela as ferramentas de TI são a infra-estrutura básica de qualquer sistema de *Business Intelligence* (REZENDE, 2003).

2.2 MINERAÇÃO DE DADOS

É a aplicação de algoritmos em busca de conhecimentos implícitos e úteis sobre dados. Nessa etapa ocorre a busca efetiva por conhecimentos úteis no contexto da aplicação de KDD.

2.2.1 Tarefas de Mineração de Dados

A Mineração de Dados é dividida, geralmente, em duas categorias: **preditivas** e **descritivas**. Algumas subdivisões são classificadas nas categorias. Sendo, a tarefa preditiva é subdividida em *classificação* e *regressão*. A tarefa descritiva em regras de *associação*, *clusterização* e *sumarização* conforme ilustrado na Figura 4:

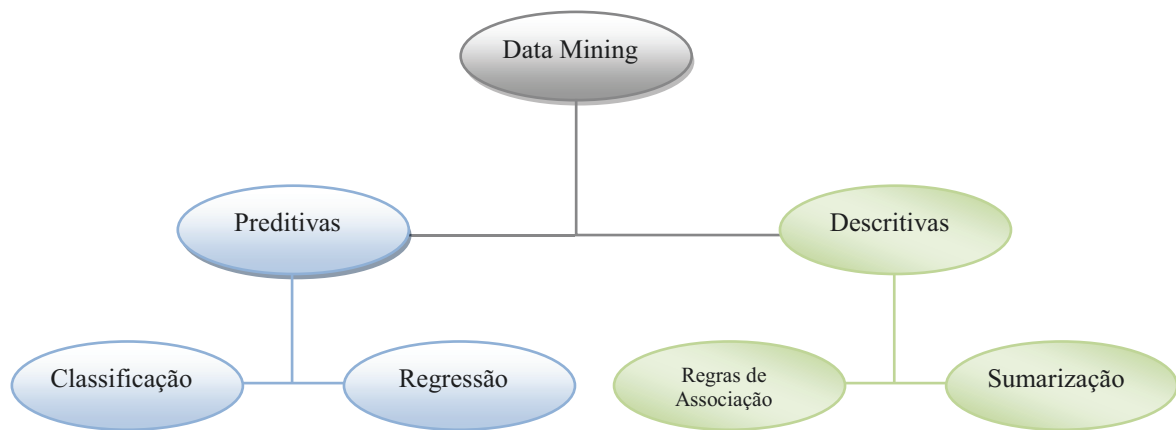


FIGURA 4 TAREFAS DE MINERAÇÃO DE DADOS
 FONTE: Adaptado de REZENDE (2003)

Tarefas Preditivas

A partir do conhecimento de alguns valores de atributos já conhecidos, as tarefas preditivas buscam identificar e/ou prever uma nova classe de atributos. O atributo a ser previsto é comumente conhecido como a **variável dependente** ou **alvo**, enquanto que os atributos usados para fazer a previsão são conhecidos como as **variáveis independentes** ou **explicativas** (TAN et.al, 2006). Se refere a tarefa de construir um modelo para a variável alvo como uma função das variáveis explicativas.

Há dois tipos de tarefas para a modelagem de previsão: classificação e regressão. A classificação prediz em classes valores alvo discretos e a regressão modela variáveis alvo em funções contínuas. As tarefas de previsão apresentam um grau de certeza em relação às variáveis alvo de valor previsto e real.

Classificação e Aprendizado

A tarefa de classificação é uma das mais importantes e populares do KDD. Esta examina características dos valores atribuídos ou uma dada situação, ou seja, associa cada atributo de uma relação a uma nova categoria de dados, uma classe. Como no exemplo da Figura 5 abaixo, em um banco de dados com registros X_i , a

função permite associá-los corretamente em uma única classe categórica Y_i , formalizados corretamente na forma ou em vetores de entradas $(x, f(x))$.

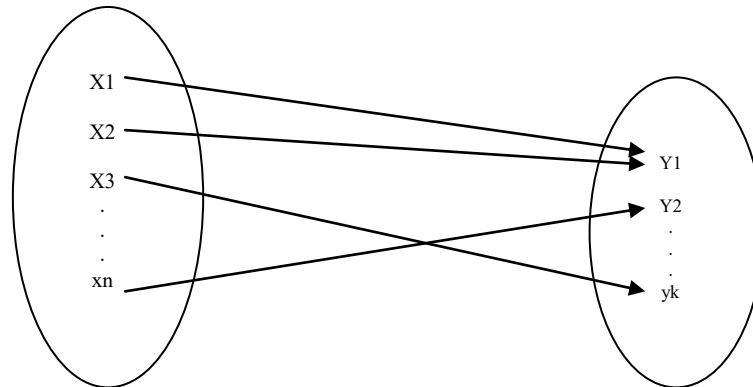


FIGURA 5 ASSOCIAÇÃO ENTRE REGISTROS DE DADOS E CLASSES.
FONTE: Adaptado de GOLDSCHMIDT e PASSOS (2006, p.67).

A tarefa de classificação é considerada preditiva, ou seja, prevê uma nova classe de dados ou em que classe um dado será envolvido a partir de uma relação de valores em conjunto de dados. Assim, surgindo a possibilidade de antecipar tendências futuras.

Num espaço de hipóteses, o processo de identificação de uma função h e mais se aproxime da função original f é chamado de **algoritmo de aprendizado** (RUSSEL, 1995 apud. cit. GOLDSCHMIDT e PASSOS, 2006). Alguns algoritmos de aprendizado aplicáveis a tarefa de classificação podem ser: C4.5, Redes Neurais, Back-Propagation, K-NN, Classificadores Bayesianos e Algoritmos Genéticos.

Na Descoberta de Conhecimento em Bases de Dados (KDD), em tarefas de Mineração como a de classificação cada algoritmo disponível pode ser utilizado ou experimentado a fim identificar qual apresenta melhor desempenho e/ou melhor resultado.

Um exemplo de classificação: Pode ser pré-selecionadas idades de um conjunto de clientes para identificar um público alvo para negócios a partir da faixa etária destes. O método apresentará o percentual relativo ao conjunto de dados disponível sobre perfis que se apresentam em idades de 10 até 17, de 18 até 25 anos, etc.

Regressão, estimativa ou predição

Similar a tarefa de classificação, segundo Goldschmidt (2006) é restrita a atributos numéricos, busca funções lineares ou não a fim de mapear registros em um banco de dados com valores reais. Essa tarefa também é conhecida como estimativa ou preditiva. Pode atribuir valor em variáveis com valores desconhecidos a partir de valores conhecidos. Considere o exemplo dado por TAN, Pan-Ning et.al, (2006): D denota um conjunto de dados que contêm N informações, $D = \{(X_i, Y_i) | i = 1, 2, \dots, N\}$. Cada X_i corresponde ao conjunto de atributos de observações de índice i (também conhecido como **variáveis explicativas**) e Y_i corresponde a **variável alvo** (ou resposta), ou seja, é a tarefa de aprender uma **função alvo** f que mapeie cada conjunto de atributos com um erro mínimo X em uma saída de valores contínuos Y

Exemplos de regressão é a previsão de: índice de bolsa de valores usando indicadores econômicos, quantidade de precipitações de uma região geográfica baseado em característica dos ventos, projeção de vendas baseada em custos com publicidade, idade de um fóssil de acordo com sua quantidade de carbono-14, valor de um imóvel de acordo com sua localidade, dimensão, segurança, etc.

Atividades descritivas

A descrição dos conjuntos de dados visa identificar e classificar atributos que não estão determinados numa classe definida para relacionamento de padrões como comportamentos semelhantes nos dados. Segundo Tan, et.al, (2006) a atividade descritiva pode servir para modelo explicativo que apresente a distinção dos dados no conjunto disposto e classes diferentes com o objetivo de derivar padrões que apresentem um breve relacionamento nos dados (correlações, tendências, grupos, trajetórias e anomalias). A seguir são apresentados três das tarefas centrais de Mineração de Dados para a modelagem descritiva: regras de associação, clusterização e sumarização.

Regras de Associação

Segundo Agrawal et al., (1993) (apud GOLDSCHMIDT, 2006) a tarefa de associação, intuitivamente, consiste em encontrar conjuntos de itens que ocorram simultaneamente. Pode ser definida formalmente como a busca por regras de **associação freqüentes e válidas**. Estuda um padrão de relacionamento entre itens de dados. Como exemplo, uma venda que ocorre simultaneamente a compra de dois produtos, páginas da web acessadas juntas, o hábito de telespectadores em assistir canais de TV de forma intercalada num mesmo horário.

A tarefa de Associação é a regra utilizada neste trabalho. Sua descrição será mais bem apresentada na seção 3.4.

Clusterização

A tarefa de clusterização é também chamada de *Agrupamento*. Para Tan, et.al, (2006) a tarefa consiste em identificar grupos de dados e dividí-los para algum significado seja para compreensão dos dados ou para utilidade apresentados em *clusters*. Para Goldschmidt (2006) esta pode ser definida como umas das tarefas básicas da Data Mining porque auxilia a realizar agrupamentos naturais de registros em um conjunto de dados. A análise de clusters organiza valores em um conjunto de padrões agrupados de acordo com algum grau de similaridade ou propriedades em comum em relações a outros grupos.

Formalmente, a clusterização supõe-se a existência de n pontos de dados X_1, X_2, \dots, X_n tais que cada ponto pertença a um espaço d dimensional R (GOLDSCHMIDT e PASSOS, 2006). Consistem em encontrar k pontos m_j em R :

$$\frac{\sum_i \min d^2(x_i, m_j)}{N}.$$

Os dados em memória de armazenagem agrupados normalmente efetuam sua tarefa de clusterização em matrizes, como no Quadro 1:

1	0	0
0	1	0
0	0	1

QUADRO 1 MATRIZ DE CLUSTER

FONTE: Adaptado de Goldschmidt e Passos (2006).

Alguns algoritmos de clusterização podem ser citados: K-Means, Fuzzy K-Means e K-Metoid.

Sumarização

A sumarização é uma tarefa utilizada para apresentação de relatórios e para análise exploratória dos dados. É também conhecida como descrição de conceitos porque apresentam de forma concisa e compreensível as principais características dos dados contidos em seu conjunto (GOLDSCHMIDT e PASSOS, 2006).

Um exemplo de sumarização pode ser a identificação de características em um conjunto de perfis de usuários de uma determinada região, com faixa salarial de X reais, nível superior incompleto e que moram em residência própria. Tais conceitos referem-se a uma classe de objetos com caracterização resumida com possível discriminação entre perfis selecionados.

3 FASES PROCESSO KDD

Para uma satisfatória operação de busca e/ou tratamento de informação com foco em utilização de ferramentas que possam lidar com diversos tipos de dados (contidos em bases, arquivos simples, planilhas ou tabelas) é necessário identificar alguns componentes envolvidos no processo de KDD: o problema a ser aplicado; os recursos disponíveis para a solução do problema e; os resultados obtidos a partir da aplicação para a solução do problema, GOLDSCHMIDT e PASSOS (2005). Desta forma, se realiza o processo em três etapas operacionais: Pré-processamento, Mineração de Dados e Pós-processamento. As etapas podem ser visualizadas na Figura 6:

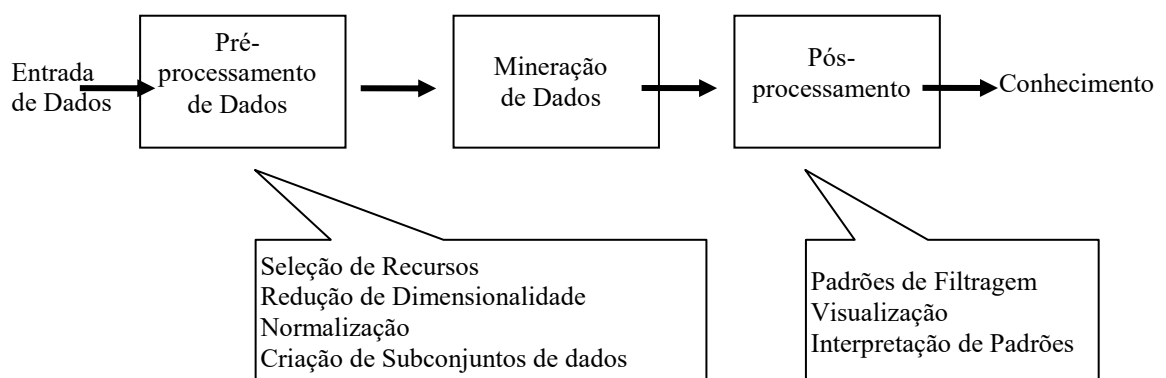


FIGURA 6 FASES DO PROCESSO DE KDD

FONTE: Adaptado de TAN, Pan-Ning et.al. (2006, p.4)

3.1 IDENTIFICANDO OS DADOS

Para que fosse conduzido o processo de Mineração de Dados foram aplicados alguns métodos orientados para a identificação de atributos que descreveriam o perfil do cliente telespectador universitário a fim de que fosse possível apresentar conjuntos relacionados à investigação e produção e resultados satisfatórios em bases de dados completos propostas por PASSOS e GOLDSCHMIDT (2005).

Após a aplicação da pesquisa no Setor Sociais Aplicadas da UFPR, foi realizado um estudo a fim de tratar atributos nos formulários recolhidos-levantados.

Os dados foram tratados para análise de perfis dos universitários como telespectadores para que a TV Band Curitiba busque apresentar programas e serviços adequados à preferência de um público alvo ou identifique uma nova oportunidade de negócio. Assim, foram identificados dados que iriam compor o escopo tratado pelo trabalho e orientariam a busca pela solução. Para ressaltar, os dados são resultados de entrevistas realizadas ao longo do ano de Nov/2009-jun/2010. Os dados foram tabulados, selecionados, transformados para valores numéricos ou nominais conforme descritos na seção de coleta de dados e outros atributos iniciais desconsiderados devido a algumas alteração no quadro de programação da Band Curitiba no ano de 2010.

3.2 COLETA DE DADOS

A fim de encontrar algumas diferenças reais entre pessoas, o questionário permitiu a pesquisa quanto a características relativamente estáveis e variáveis nos indivíduos. Para Selltiz (1974), devem ser abordados no questionário características variáveis como formação, inteligência e personalidade das pessoas porque são relativamente estáveis no indivíduo e podem influenciar ou contaminar resultados. Estas devem estabelecer ligação com o problema e os objetivos da pesquisa, a população a serem pesquisados, os métodos de análise escolhidos e/ou disponíveis.

Uma forma de investigação em princípios de levantamentos de atividades em função dos objetivos com algumas ações técnicas a serem verificadas antes do processo de Mineração de Dados com a utilização de instrumentos como um software (GONÇALVES, 2001), podendo ser selecionadas algumas tarefas para formação de uma base de dados, para a qual será realizado este trabalho.

Os dados foram coletados no Campus Botânico da Universidade Federal do Paraná especificamente no Setor Ciências Sociais Aplicadas afim de criar um conjunto mínimo necessário para a classificação de perfis.

Os atributos relativos aos de telespectadores universitários são de perfil acadêmico, econômico, social e de marketing empresarial. Os dados foram armazenados em planilhas dados da Microsoft Excel e convertidos em arquivo para o formato arff para processamento na interface Weka, contendo 300 estudantes. Os

campos utilizados para bases e cálculo de cada valor descrito como resultantes de fórmulas ou indicadores como: curso, turno, período ou estágio de graduação, sexo, idade, estado civil, convívio familiar, faixa salarial, naturalidade e cidade de residência.

3.3 PRÉ PROCESSAMENTO

Compreende todas as funções relacionadas à captação, à organização e ao tratamento e à preparação dos dados para os algoritmos da etapa da Mineração de Dados. Essa etapa realiza o principal processo de Descoberta de Conhecimento desde a correção de dados errados até o ajuste da formatação dos dados a serem utilizados.

Os campos ou atributos a serem utilizados para cálculo dos indicadores são: horários, preferência por apresentadores locais e programação da Band Curitiba. Estas foram devidamente especificadas durante a coleta de dados a fim de simplificar a sequência de treinamento dos algoritmos de mineração na base Weka. No arquivo, foram encontrados alguns valores ausentes como a data de nascimento e período de grau de graduação. Os valores foram ocupados por valores nulos 0 (zero) para completar a base, assim indicando período não informado. Na pesquisa, em algumas situações com a pesquisadora presente no momento do preenchimento, puderam-se esclarecer dúvidas dos entrevistados quanto à revelação da idade. Reação comum entre jovens, principalmente do sexo feminino.

Atributos que descrevem a posição do entrevistado com opinião referente a questões como preferência pela programação da TV Band Curitiba foram descartados apenas da base para processamento no Weka, pois deixariam a base de dados com informações inconsistentes. Para orientação e organização de dados o trabalho foi realizado em planilhas do software Microsoft Office Excel e tabelas em Microsoft Office Word para posteriormente serem transferidos para texto puro em bloco de notas em formato Arff.

Desta forma as técnicas de pré-processamento utilizadas para a pré-seleção dos dados foram definidas considerando a natureza dos dados por suas características a seguir, como na Tabela 1:

TABELA 1 CLASSIFICAÇÃO DAS VARIÁVEIS NO CONJUNTO DE DADOS

VARIÁVEL		VALOR	VARIÁVEL		CATEGORIA
Estado civil	Casado	1	Idade	< 19 anos	A
	Solteiro	2		Entre 20 e 24 anos	B
	Viúvo	3		Entre 25 e 29 anos	C

FONTE: Adaptado de (GOLDSCHMIDT e PASSOS 2006).

- Variáveis de categoria – apresentam valores de um conjunto de possibilidades limitado e pré-determinado;
- Variáveis Numéricas – permite operações de aritméticas e de ordenação.

3.4 TÉCNICA DE ASSOCIAÇÃO UTILIZADA

A localização de padrões visa identificar seqüências de informações de ocorrência repetitiva. Isto não é uma tarefa privilegiada ao Data Mining, intuitivamente a capacidade humana é capaz de assimilar uma dada freqüência. Mas, em grandes bases de dados ou em uma amostra significativamente volumosa a manutenção humana convencional é incapaz de encontrar. Essa tarefa foi introduzida por Agrawal et al., (1993) (apud GOLDSCHMIDT, 2006) afirmando que, intuitivamente, a tarefa associativa consiste em encontrar conjuntos de itens que ocorram simultaneamente. Pode ser definida formalmente como a busca por regras de **associação frequentes e válidas**. Estuda um padrão de relacionamento entre itens de dados.

As etapas da tarefa associativa consistem em perceber que existe uma seqüência de ocorrências que se repetem; a próxima etapa é permitir que o analista verifique se ocorrem com freqüência superior às demais freqüências; confirmar se elas segmentam o padrão original e independente; finalmente, fazer induções, gerar representações genéricas das ocorrências.

As representações genéricas substituem os dados brutos em expressões simples. Se um conjunto de itens é freqüente, então todos os seus subconjuntos são freqüentes, (TAN, P. et.al, 2006).

Em tarefas de associação alguns termos são utilizados na aplicação de sua forma. Segundo Goldschmidt (2006) a *transação* numa associação é o nome atribuído ao elemento de ligação existente em cada ocorrência em dados; a associação é considerada *frequente* se o número de vezes em que a união de um conjunto de itens for superior, em relação ao total de transações, a uma frequência mínima estabelecida em cada aplicação; a associação é considerada *válida* se a freqüência ocorrer com um valor denominado *confiança mínima*. Assim, a associação é especificada a partir de parâmetros de suporte e confiança mínimos. Mais importante do que simplesmente obter essa redução (compressão) de informação, as tarefas de associação nos permite gerar formas de *predizer* futuras ocorrências de padrões. Este é exatamente o ponto onde é apresentado um resultado de valor.

3.4.1 O Algoritmo Apriori

Quando o assunto é regras de associação em conjuntos de dados, a técnica mais conhecida e utilizada neste sentido é o algoritmo *Apriori*, um algoritmo clássico de Mineração de Dados (AGRAWAL, 1993). O princípio diz que “um *k-itemset* somente pode ser freqüente se todos os seus *(k-1)-itemsets* forem frequentes”. Com a técnica é possível encontrar todas as ocorrências frequentes denominadas de itens, ou *itemsets frequentes* (L_k).

O primeiro passo do algoritmo é encontrar itens frequentes, ou seja, realizar a contagem de ocorrências dos itens a partir da condição especificada no parâmetro de suporte mínimo, determinar seu tamanho unitário e eliminar itens que não são frequentes. O segundo passo extrai as sequencias dividindo-se em duas fases: (1) gerar itens potencialmente frequentes, *itemsets* candidatos C_k ; (2) o conjunto de dados é processado novamente realizando uma busca contando candidatos pelo seu parâmetro de suporte. Então, satisfazendo a condição de suporte mínimo. Assim, L_{k-1} é o conjunto de todos $(k-1)$ satisfazendo a condição de *itemsets*

candidatos frequentes, ou seja, se um *itemset* X tem suporte mínimo, todos os seus subconjuntos também terão (AGRAWAL, 1993).

Finalmente, a partir do conjunto de dados com itens frequentes são descobertas as regras de associação na condição de confiança mínima, significa encontrar todos os *subsets* produzindo uma regra *a*, se esta é ao menos igual à confiança mínima gerando conseqüentes denominados *subsets*. A Figura 7 ilustra as regras geradas:

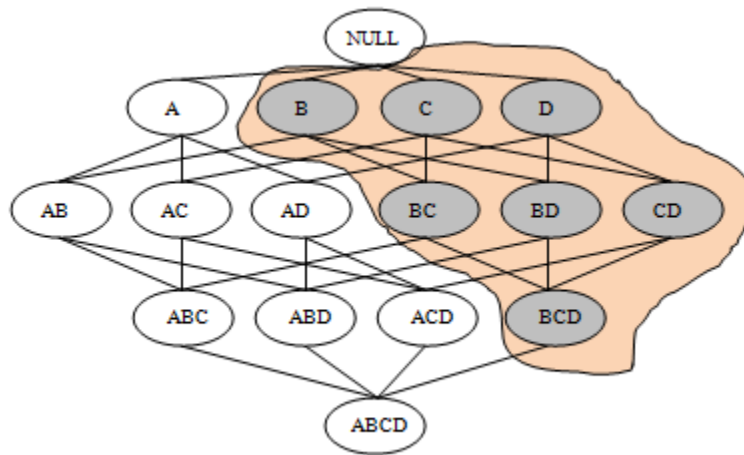


FIGURA 7 EXEMPLO DE CONJUNTOS DE ITENS FREQUENTES
FONTE: AUTORA Adaptado de TAN et. al (2006)

Os parâmetros de suporte e confiança são determinados pelo usuário. Suporte é uma técnica que reduz o número de itens e combinações a serem considerados nas transações, ou seja, elimina itens que não aparecem em número suficiente nas transações (BERRY e LINOFF, 1997). Confiança é a relação entre o número de transações entre todos os itens e os itens válidos gerando a regra: „SE“ *B* e *C* „ENTÃO“ *A*. Mede a relação de itens que fazem parte do „SE“, medindo o quanto um item depende do outro, ou seja, numa frequência da transação *B* e *C* diz a porcentagem de que *A* também apareça.

3.5 PÓS-PROCESSAMENTO

Abrange o tratamento do conhecimento obtido na Mineração de Dados. Trata-se de facilitar a interpretação e a avaliação da utilidade do conhecimento descoberto, sendo: elaboração e organização, podendo incluir a simplificação a partir de gráficos, diagramas, ou relatórios demonstrativos; além da conversão da forma de representação do conhecimento obtido.

É na etapa do pós-processamento que resultados das aplicações de Mineração de Dados são avaliados e novas alternativas para identificação dos dados são compreendidas como conhecimento obtido do processo de KDD. Este é validado e representado de diversas formas conforme já mencionado.

Goldschmidt e Passos (2006) apresentam duas operações de pós-processamento: a *Simplificações de Modelo de Conhecimento* que remove detalhes a fim de a representação torne-se menos complexa para compreensão e considerando toda sua relevância. E *Transformação de Modelo de Conhecimento* que a partir de modelos utilizados realiza métodos de transformação nestes, ou seja, uma forma diferente de representação do mesmo modelo apenas para facilitar a análise.

4 METODOLOGIA

Neste item são apresentados o delineamento das variáveis, população e amostragem, coleta e análise dos dados e a metodologia utilizada para processamento dos dados, ou seja, a Ferramenta Weka. A pesquisa é tratada neste capítulo com a abordagem de características pessoais, perfil social, econômico e nível de interação com a TV dos estudantes universitários do Setor de Ciências Sociais Aplicadas da UFPR.

4.1 DELINEAMENTO E DELIMITAÇÃO DA PESQUISA

A pesquisa é caracterizada como pesquisa quantitativa tendo como ambiente um canal de televisão aberta – TV Band e como amostra os alunos dos cursos de Administração, Economia, Contabilidade e Gestão da Informação, da Universidade Federal do Paraná, Setor de Ciências Sociais Aplicadas.

Quanto à abordagem quantitativa significa traduzir em números opiniões e informações para classificá-los e analisá-los. Neste trabalho, foi empregada a pesquisa descritiva. Trata-se de um levantamento realizado para quantificar e tornar relevante alguns aspectos interessantes para empresas. Segundo Malhotra (2001) um tipo de pesquisa que tem como principal objetivo a descrição de algo – normalmente características ou funções de mercado. Trata-se de uma concepção básica de pesquisa gerada a partir do conhecimento prévio em relação à situação-problema, definindo claramente as informações necessárias pré-planejadas e estruturas se baseando em amostras representativas pelo método de coleta de dados sobre as mesmas.

4.2 AMBIENTE DO ESTUDO DE CASO

Aqui é detalhado o processo de regras de associação, ou seja, a descrição de dados a partir de padrões de valores nas ocorrências frequentes até chegar ao conhecimento desejado no ambiente escolhido para a pesquisa.

As empresas de TV aberta estão atentas para modificações necessárias no tocante a acirrada concorrência regional e até nacional no ramo de comunicação que vieram com a tecnologia. Assim, surge o aumento da importância dada à informação que atinge clientes em forma tanto tangível (com produtos impressos), quanto intangível (prestação de serviços personalizados, pessoais, e hoje, até de forma virtual – pela Internet). Torna-se necessária que a informação evite a implicação de riscos e ofereça oportunidades. A execução de algoritmos de Mineração de Dados explora conjuntos de dados de forma a propiciar a descoberta ou produza modelos de conhecimento (GOLDSCHMIDT e PASSOS, 2005).

Para garantir sustentabilidade no mercado, a organização deve dispor de uma base com dados primordiais referente ao seu negócio – a de seus clientes. Sendo, exercer controle sobre informações de seus usuários com competência tecnológica, irá garantir um diferencial estratégico com relacionamento e satisfação e/ou reciprocidade na prestação de serviços – no caso uma programação de TV. Isso permitirá tanto para o cliente uma seleção e identificação de serviços adequados às suas exigências, quanto para a empresa que oferece seu produto de acordo com tais perfis descobertos em regras ou ocorrências padronizadas nos dados armazenados. Surgindo, desta forma, maior possibilidade de conquista e retenção de seus usuários.

Assim, este trabalho oferece um conjunto de dados identificados e analisados sobre clientes universitários para TV a partir da Descoberta de Conhecimento em Bases de Dados. Isso tornou uma base propícia à descrição de eventos e a associação de dados sobre telespectadores para métodos de Mineração de Dados, possibilitando visões mercadológicas tais como análises de fatias de mercado, padrões de consumo de consumidores, estudos de seus comportamentos, estilos de vida, opinião, informação para investimento em propaganda e medidas de audiência em canais de televisão.

Tal pesquisa busca tratar o problema da TV Band Curitiba em manter a fidelidade do telespectador, programação inadequadas ou recusadas pelos clientes e trazer soluções para apresentação do produto ou serviço com conteúdo e horário personalizado às exigências a ela atribuídas.

O objetivo deste estudo de caso é comprovar a adequação da metodologia de regras de associação na identificação de padrões usando o algoritmo Apriori disponível na ferramenta *Waikato Environment Knowledge Analysis - Weka*.

TABELA 1: UMA CONCEPÇÃO BÁSICA DE PESQUISA

Tipologia	<i>Descritiva</i>
Objetivo:	Descrever características ou funções do mercado
Características :	Marcada pela formação prévia de hipóteses específicas Concepção pré-planejada e estruturada
Métodos	Dados Secundários Levantamentos Painéis Dados de observação e outros dados

FONTE: AUTORA COM BASE NA DESCRIÇÃO DE MALHOTRA (2004)

A pesquisa descritiva usa métodos de aplicação que podem ser classificados como longitudinais e transversais. Neste trabalho foi empregado o método de estudo transversal, envolve a coleta de informações de qualquer amostra de elementos da população somente uma vez, também conhecida por concepção de pesquisa por levantamento de amostragem, (MALHOTRA, 2004).

4.3 APRESENTAÇÃO DAS VARIÁVEIS

Segundo Kerlinger (1980), variável “é um constructo, um conceito com um significado especificado „construído” dado por um pesquisador”. Um termo variável pode ser constituído a partir de definição de “palavras com outras palavras” possíveis de serem encontradas em dicionários e usadas por qualquer pessoa, inclusive pesquisadores. Com relação às características econômicas e sociais a classificação seguiu as categorias utilizadas pelo IBGE.

As variáveis da pesquisa foram:

- a) características acadêmicas do universitário (nome do curso, grau e horário acadêmico);
- b) características pessoais (familiaridade, gênero, renda e residência);

- c) características pessoais de lazer ou organização referente à disponibilidade aos serviços de mídia local oferecidos pelas TVs abertas (horários e preferências);
- d) avaliação da preferência ou conhecimento da programação apresentada pela TV Band Curitiba.

As definições constitutivas das variáveis e sua codificação para tratamento dos dados são apresentadas no item Sistematização e Análise dos Dados.

4.4 POPULAÇÃO E AMOSTRAGEM

Segundo Malhotra (2001), população “é o agregado, ou soma, de todos os elementos que compartilham algum conjunto de características comuns, conformando o universo para o problema de pesquisa de marketing” (p. 301), “é um subgrupo de uma população, selecionado para participação no estudo” (p. 301).

A população total é constituída por 3.300 alunos, sendo 2.100 alunos da graduação. Portanto, foram selecionados 300 alunos o que representa 14,28% da população de alunos de graduação do respectivo setor. A escolha dessa população-alvo deve-se à proximidade da pesquisadora no campus, pela possibilidade de universitários representarem uma classe social e um público definido para um segmento de mercado com potencial consumo e/ou utilização dos serviços de TV.

A pesquisa é baseada em amostra não aleatória do tipo **intencional**: composta por elementos da população selecionados intencionalmente pelo investigador, (MARCONI & LAKATOS, 1996; OLIVEIRA, 1997). O investigador pode considerar esses elementos de características típicas ou representativa da população, como: a opinião de um público definido de universitários para a TV aberta local. Assim, foram selecionados, intencionalmente, 300 alunos dos cursos de graduação do Setor de Ciências Sociais Aplicadas, conforme apresenta o Gráfico 1:.

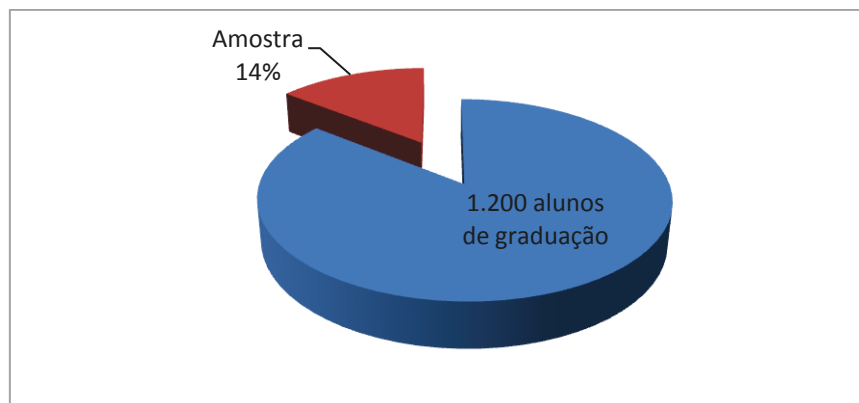


GRÁFICO 1 POPULAÇÃO E AMOSTRAGEM
FONTE: AUTORA COM BASE NOS DADOS DA PESQUISA

A técnica de amostragem empregada foi por conveniência que de acordo com Malhotra (2001) utiliza seleção aleatória, confiando no julgamento do pesquisador. A conveniência visa alcançar uma amostra de elementos convenientes, nesta pesquisa, a disponibilidade do universitário em relação à TV e à sua programação local.

4.5 COLETA DE DADOS

Os dados foram coletados a partir de um questionário elaborado com perguntas fechadas para identificação de perfis estudantes de qualquer faixa etária, independente da renda familiar, conhecimento ou disponibilidade definida para TV conforme o interesse da empresa integrada no estudo de caso. Na aplicação dos questionários, procurou-se obter um equilíbrio entre a amostra nos turnos matutino e noturno, no intuito de evitar algum erro ou tendenciosidade.

Algumas questões iniciais do questionário elaborado em 2009 foram modificadas devido a alterações na Programação do Canal Aberto de Televisão – TV Band de forma a enquadrar os atributos relativos aos interesses passíveis de serem identificados em conjunto de dados e estruturados para Mineração de Dados.

O questionário foi dividido em quatro seções: a primeira seção teve como objetivo identificar características acadêmicas do universitário; a segunda seção identifica características pessoais do entrevistado; depois, a identificação da preferência do universitário em relação à TV aberta local; e a quarta seção permite que o estudante indique sua(s) preferência(s) com relação à programação da Band

Curitiba. Desta forma, os dados relativos à frequência de usuários preferenciais a partir do mês de novembro de 2009 a junho de 2010 no canal da Band Curitiba com pesquisa neste sentido mediante preenchimento de formulário, conforme apresentado no Apêndice A.

Os selecionados foram 300 estudantes preferencialmente de último e penúltimo ano de conclusão de curso. Alguns entrevistados de outros períodos foram abordados devido à dificuldade de disponibilização e desperiodização em conjunto nas salas de último grau de graduação. O propósito da pesquisa foi distinguir grupos de estudantes por curso e turnos para avaliação do objeto de pesquisa para TV aberta local.

A coleta de dados pode ser definida como uma aplicação objetiva, que é informada uma única vez pelo usuário permitindo respostas a questões simples, ou seja, em perguntas fechadas que apresentam questões em categorias ou alternativas de respostas fixas e em perguntas abertas para conhecimento da empresa interessada na pesquisa. O levantamento de dados é o principal componente do processo e consiste na coleta de dados com objetivo de delinear o perfil do usuário telespectador abrangendo, em algumas considerações, as preferências da família e comunidade nele envolvido, ou seja, a pesquisa em seu sentido de identificação de preferências de telespectadores através da prática descritiva a partir de seus perfis.

Para compreender o termo perfil, considera-se a descrição básica a partir do dicionário como Michaelis (2009) que considera o termo *perfil* como:

*1 Contorno ou delineamento do rosto de uma pessoa, visto de lado.
6 Pequeno escrito em que se salientam os traços característicos de uma pessoa. _P. psicológico: gráfico que representa as diversas aptidões de uma pessoa. De perfil: de lado.*

Haveria, então, uma correlação entre o termo *perfil* e o termo *paradigma*? Para *paradigma* segundo Michaelis (2009):

1 Modelo, padrão, protótipo. 2 Ling Conjunto de unidades suscetíveis de aparecerem num mesmo contexto, sendo, portanto, comutáveis e mutuamente exclusivas. No paradigma, as unidades

têm, pelo menos, um traço em comum (a forma, o valor ou ambos) que as relaciona, formando conjuntos abertos ou fechados, segundo a natureza das unidades.

TABELA 2 DESCRIÇÃO DE PERFIL E PARADIGMA

PERFIL	<i>Descrição básica e concisa</i>
PARADIGMA	<i>Modelo, padrão</i>

FONTE: AUTORA COM BASE NA DESCRIÇÃO UOL. Michaelis Moderno Dicionário (2009)

Portanto, neste trabalho, deve-se, descrever os elementos identificadores das características básicas do telespectador universitário, enquanto o *perfil* procura delinear o objeto de análise a partir de sua singularidade e sem interferências externas. Considerando-se que o termo paradigma significa *modelo, padrão*, pode-se inferir que o mesmo seja a cristalização do protótipo em seu aspecto referencial, para que o objeto de análise deixe de ser tão somente um resultado para servir de base ou guia para um raciocínio indutivo, do específico (modelo) para o geral. Como se tratando aqui de uma categoria ligada à idéia do perfil do universitário do Setor de Ciências Sociais Aplicadas da UFPR em suas divisões internas como: estudantes dos cursos de Administração, Economia, Contabilidade e Gestão da Informação.

4.6 SISTEMATIZAÇÃO E ANÁLISE DOS DADOS

A sistematização e análise dos dados foram realizadas segundo as definições constitutivas para definições operacionais das variáveis estabelecidas para atribuir significado a uma variável, determinando uma maneira de medi-la ou manipulá-la. As variáveis são apresentadas a seguir:

Para Características acadêmicas foram definidos:

Graduação

DC: “Divisão do círculo em graus, minutos e segundos; posição social” (UOL MICHAELIS).

DO: **Curso** - Valor criado para identificar a graduação do universitário. Os valores foram atribuídos para Administração (1); Contabilidade (2); Economia (3); Gestão da Informação (4) e; (0) para curso não informado.

DO: **Turno** - Valor criado para identificar o horário da principal atividade acadêmica do entrevistado. Os atributos são para manhã (1); noturno (2) e; (0) para turno não informado.

DO: **Período** - Valor de atributo numérico criado para identificar o grau de graduação do estudante. Para cursos com períodos anuais foram calculados como em divisão semestral para simplificar a tabulação. Por exemplo: 2º ano = 4º período.

Características pessoais

Relativo ao termo pessoal, define-se:

DC: “Pertencente ou relativo à pessoa; que é próprio e particular de cada pessoa; exclusivo de certa pessoa; individual.” (UOL MICHAELIS).

As características pessoais foram operacionalizadas por meio de outras variáveis: gênero, familiaridade, conhecimento prévio, renda e preferência.

Gênero

DC: “Agrupamento de indivíduos que possuem caracteres comuns” (UOL MICHAELIS).

DO: **Sexo** - valor criado para identificar o gênero do entrevistado. Os valores atribuídos são para masculino (1) e feminino (2).

Idade

DC: “Número de anos de alguém ou de alguma coisa” (UOL MICHAELIS).

DO: **Idade** - Classificado em variáveis de categoria por faixa etária: A, B, C e D.

Familiaridade

DC: “Familiaridade é qualidade daquele ou daquilo que é familiar; Confiança; Franqueza; Intimidade” (UOL MICHAELIS).

DO: **Estado_civil** - Valor criado para identificar o convívio familiar e/ou social do estudante. Os valores foram atribuídos para se casado (1); solteiro (2); viúvo (3); separado (4) e; (5) para união estável.

DO: **Convívio familiar ou pessoas do convívio** - Valores criados para complementar a questão anterior identificando o as pessoas do convívio envolvido. Os valores foram separados para simplificar a tabulação sendo: convívio1 (marido/companheiro(a); 2 (filhos); 3 (só); 4 (pais) e; 5 (outros) este atributo foi excluído da tabulação devido a descrição aleatória fornecidas pelos entrevistados.

Renda

DC: “Importância superior a determinado limite e estabelecida pelas leis fiscais como rendimento da atividade econômica do indivíduo” (UOL MICHAELIS).

DO: **Faixa_salarial** - Valor criado para identificar o nível econômico do estudante. Os valores em classes foram: até R\$500 (1); de R\$500 a R\$1.000 (2); de R\$1.000 a R\$2.000 (3); de R\$2.000 a R\$3.000 (4) e; acima de R\$3.000 (5).

Naturalidade

DC: “A pátria; Naturalização. *A n. de alguém:* a sua pátria; a terra onde nasceu” (UOL MICHAELIS).

DO: Valor criado para identificar a região natal do estudante. As cidades, estados e países foram numerados conforme a sequencia de revisão dos formulários. Para cidades fora do estado do Paraná foram classificadas as região indicando mesorregiões para simplificar a operação. Por exemplo: Curitiba = Capital; Londrina (Norte Central).

- | | | |
|----------------------|--------------------------|------------------------|
| 1. Curitiba | 18. Santo Antonio da | 34. Umuarama |
| 2. Campo Largo | Platina | 35. Cascavel |
| 3. Campo Mourão | 19. Santa Catarina | 36. Avellaneda – |
| 4. Ponta grossa | 20. Piauí | Argentina |
| 5. Pato Branco | 21. Maceió | 37. Sarandi |
| 6. Paranaguá | 22. Arapongas | 38. Miami |
| 7. São Mateus do Sul | 23. Três Barras | 39. Faxinal |
| 8. Lapa | 24. Rio de Janeiro | 40. Sertanopolis |
| 9. Alm. Tamandare | 25. Dois Vizinhos | 41. Minas Gerais |
| 10. Londrina | 26. Paranavaí | 42. Guarapuava |
| 11. Dourados | 27. Foz do Iguaçu | 43. Rio Grande do Sul |
| 12. Curupi TO | 28. Mato Grosso do Sul | 44. Jandaia do Sul |
| 13. Vilhena RO | 29. São Jose dos Pinhais | 45. Pinhais |
| 14. Salvador | 30. Safragma | 46. Fazenda Rio Grande |
| 15. Ibiporã | 31. Quitandinha | 47. Cambara |
| 16. São Paulo | 32. Colombo | 48. Muqui - ES |
| 17. Maringá | 33. Jaguariaíva | |
| 49. | | |

Cidade_res

Valor criado para identificar o local/cidade da residência do estudante. A identificação dos atributos seguiu a numeração apresentada na questão anterior.

Seção de identificação da preferência do universitário em relação à TV aberta local

Opinião, Conhecimento e Preferência

DC: Opinião - “Maneira de opinar; modo de ver pessoal; parecer, voto emitido ou manifestado sobre certo assunto” (UOL MICHAELIS).

DC: Conhecimento - “Idéia, noção; informação, notícia.; *de causa*: perícia; experiência” (UOL MICHAELIS).

DC: Preferência - “Ação ou efeito de preferir uma pessoa ou coisa a outra; predileção” (UOL MICHAELIS).

DO: estas variáveis foram mensuradas para apontamentos que indicavam experiência, procura e uso da TV aberta local. Foram utilizados valores que confirmavam o uso da TV com a opção de seleção de horários. Para canais específicos, a opção de seleção do apresentador foi um indicador de referencia profissional ou canal de maior experiência.

Para horários em que o universitário costuma assistir a programação local foram separados em atributos para simplificar a análise:

- *Horario1* (das 6 às 7h).
- *Horario2* (das 7 às 8h).
- *Horario3* (das 8 às 9h).
- *Horario4* (das 9 às 10h).
- *Horario5* (das 10 às 11h).
- *Horario6* (das 11 às 12h).
- *Horario7* (das 12 às 13h).
- *Horario8* (das 13 às 14h).
- *Horario9* (das 14 às 15h).
- *Horario10* (das 15 às 16h).
- *Horario11* (das 16 às 17h).
- *Horario12* (das 17 às 18h).
- *Horario13* (das 18 às 19h).
- *Horario14* (das 19 às 20h).
- *Horario15* (das 20 às 21h).
- *Horario17* (das 21 às 22h).
- *Horario18* (das 22 às 23h).
- *Horario19* (das 23 às 24h).
- *Horario20* (a partir das 0h).

Para apresentação e/ou programas de preferência. Os atributos que não foram selecionados não estão apresentados:

- | | |
|--|---|
| 1. Adriana Milczewsky Rendak (Mapa Tempo - Bom Dia Paraná – RPC) | 4. Ana Carolina Oleksy (Previsão do Tempo – RPC) |
| 2. Alessandra Consoli (Bom Dia Paraná – RPC TV) | 5. Antônio Carlos da Silva (Carros & Motores – TV Iguaçu) |
| 3. Alphonse Voigt e Carol Facchina (Jogo Aberto Paraná - Band) | 6. Camila Paschoal (Jornal da Educativa– TVE PR) |

- | | |
|--|---|
| 7. Cecília Comel e Lia Lago (Destaque – TV Iguaçu) | 23. Odilon Araújo e Paola Manfroi (Revista RPC) |
| 8. Cesar Camargo e por Kelly (TransOfertas – TV Transamérica) | 24. Paulo Roberto (Boa Tarde PR - Band) |
| 9. Denian Couto (Jornal da Massa – TV Iguaçu) | 25. Pe. Reginaldo Manzotti (Palavras de Fé TV Iguaçu e Band Curitiba) |
| 10. Dulcinéia Novaes (Meu Paraná – RPC) | 26. Raphael dos Anjos (Paraná no Ar - RIC TV) |
| 11. Elisia Munareto (Enfoque – TVE PR) | 27. Raphael Laffranchi e Roberta Pitarelli (RIC Notícias) |
| 12. Eugênio Stefanello e Lauren Trilha (Negócios da Terra – TV Iguaçu) | 28. Roberto Accioli (190 – CNT) |
| 13. Fabrício Binder (Band Cidade e PR Acontece) | 29. Robson Silva (Tribuna da Massa 2ª Ed – TV Iguaçu) |
| 14. Gil Rocha (Revista RPC) | 30. RR Soares (Show da Fé – TV Iguaçu) |
| 15. Giselle Hishida (Band Cidade) | 31. Sandro Ivanowski (Caminhos do Campo – RPC) |
| 16. Giselle Macedo (Nossa Tarde - Band) | 32. Sérgio Mendes e Rose Machado (RIC Rural) |
| 17. Herivelto Oliveira (ParanáTV – RPC TV) | 33. Simone Munhoz (RIC Notícias) |
| 18. Janine Furtado (Plug – RPC) | 34. Thays Beleze (ParanáTV – RPC TV) |
| 19. Jasson Goulart (Globo Esporte - RPC) | 35. Valdecir Santos (Tribuna da Massa 1ªEd – TV Iguaçu) |
| 20. Lígia Gabrielli (Jornal Mercosul – Rede Mercosul) | 36. Viviane Gottardi (Vitrine Mercosul – Rede Mercosul) |
| 21. Luiz Carlos da Rocha (Jogo do Poder – CNT) | 37. Wiliam e Renan (TV Educativa) |
| 22. Mira Graçano (ParanáTV – RPC TV) | |

Assunto(s) de preferência do entrevistado com relação à programação da TV Band Curitiba.

O programa Nossa Tarde não é mais apresentado por isso foi desclassificado da tabulação, mantendo assim os programas mais consolidados:

BAND CIDADE (noticiário de 2ª a sábado às 18h50min)

- | | |
|--------------------------------------|--|
| • <i>Band_cidade1</i> (política) | • <i>Band_cidade6</i> (educação) |
| • <i>Band_cidade2</i> (vida pública) | • <i>Band_cidade7</i> (segurança) |
| • <i>Band_cidade3</i> (cultura) | • <i>Band_cidade8</i> (opção 'nunca assisti') |
| • <i>Band_cidade4</i> (esporte) | • <i>Band_cidade10</i> (valor atribuído pela sessão mantida em branco pelo entrevistado) |
| • <i>Band_cidade5</i> (saúde) | |

PR ACONTECE (noticiário de 2ª a sexta às 6h50min)

- | | |
|--------------------------------------|--|
| • <i>Pr_acontece1</i> (política) | • <i>Pr_acontece7</i> (segurança) |
| • <i>Pr_acontece2</i> (vida pública) | • <i>Pr_acontece8</i> (opção 'nunca assisti') |
| • <i>Pr_acontece3</i> (cultura) | • <i>Pr_acontece10</i> (valor atribuído pela sessão mantida em branco pelo entrevistado) |
| • <i>Pr_acontece4</i> (esporte) | |
| • <i>Pr_acontece5</i> (saúde) | |
| • <i>Pr_acontece6</i> (educação) | |

BOA TARDE PR (telejornal de 2ª a 6ª-feira às 13h30min)

- | | |
|-----------------------------------|--|
| • <i>Boa_tarde_pr1</i> (policial) | • <i>Boa_tarde_pr2</i> (solidariedade) |
|-----------------------------------|--|

- *Boa_tarde_pr3 (Quadro Lig Band)*
- *Boa_tarde_pr4 (Quadro Me Chama Que Eu Vou)*
- *Boa_tarde_pr5 (opção 'nunca assisti')*
- *Boa_tarde_pr10 (valor atribuído pela sessão mantida em branco pelo entrevistado)*

JOGO ABERTO PR (programa de esporte de 2ª a 6ª-feira às 12h30min)

- *Jogo_aberto_pr1(Quadro Rodada e Classificação)*
- *Jogo_aberto_pr2 (Quadro Que Resenha é Essa?)*
- *Jogo_aberto_pr3 (Recados)*
- *Jogo_aberto_pr4 (Comentários)*
- *Jogo_aberto_pr5 (opção 'nunca assisti')*
- *Jogo_aberto_pr10 (valor atribuído pela sessão mantida em branco pelo entrevistado)*

BAND PÉ NO RIO (programa universitário apresentado aos sábados às 19h15min)

- *Pe_no_rio1 (descrição geral sobre o programa)*
- *Pe_no_rio8 (opção 'nunca assisti')*
- *Pe_no_rio10 (valor atribuído pela sessão mantida em branco pelo entrevistado)*

O programa Band Pé no Rio não é mais apresentado, porém, devido ao considerável conhecimento pelos entrevistados, o atributo foi mantido para tabulação.

4.7 FERRAMENTA WEKA

A Ferramenta Weka foi utilizada para o processamento dos dados. A escolha desta ferramenta foi feita tendo em vista sua disponibilização gratuita e sua aderência ao objeto desta pesquisa.

Weka é um software que contém uma coleção de algoritmos para tarefas de mineração de dados. Oferece ferramentas para o pré-processamento de dados, classificação, regressão, clustering, regras de associação e visualização de dados e modelos baseados em gráficos. Foi desenvolvido por uma equipe de pesquisadores da Universidade de *Waikato* e disponibilizado gratuitamente para tanto para utilização quanto para possíveis alterações e revenda. Feito em código Java, o qual está disponível na Internet no *website* (<http://www.cs.waikato.ac.nz/~ml/weka/>), o que permite que o software seja utilizado em (possivelmente) qualquer plataforma computacional.

O sistema lê arquivos de dados armazenados somente em um formato específico chamado **arff**, desenvolvido pela equipe da Universidade de *Waikato* e contém a seguinte estrutura:

Relação, linha que indica o assunto ou a tarefa sendo representada pelos dados;

sintaxe: @relation <nome da relação>

Atributo, conjunto de linhas da relação que informa o nome do atributo e os valores/tipo (neste caso as alternativas devem ser redigidas em uma lista separada por vírgulas e cercadas por chaves), numérico (neste caso o nome deve ser seguido da palavra-chave real), data, booleano e outros.

sintaxe: <nome do atributo> <tipo ou conjunto de atributos>

Dados, apresenta-se em linhas informando cada valor individualmente correspondendo a uma instancia e com valores separados por vírgulas e na mesma ordem dos *Atributos*.

sintaxe: @data (seguido dos valores nas linhas abaixo)

O usuário do Weka deve realizar a conversão dos dados manualmente para sua estrutura em formato arff.

Para executar as tarefas de *Data Mining* no Weka é possível utilizar um conjunto de pacotes, que são: *attribute selection* (seleção de atributos de uma base de dados), *classifiers* (implementação de algoritmos de aprendizagem supervisionada), *clustering* (implementação de algoritmos de aprendizagem não supervisionada) e *filters* (seleção de instâncias de uma base). (Witten, 2005)

5 RESULTADOS: ANÁLISE E DISCUSSÃO

Neste item são apresentados: a análise e discussão os resultados obtidos por meio dos questionários e da aplicação da mineração de dados com a ferramenta Weka.

5.1 RESULTADOS OBTIDOS PELOS QUESTIONÁRIOS

O elemento chave para um perfil é a coleta e análise de dados que constitui o fundamento para a identificação de necessidades, respostas e problemas individuais. É nessa fase que se faz uma investigação sistemática e deliberada de dados para a determinação do estado atual e (talvez) passado do usuário da TV aberta, seu estado descritivo e avaliação padrão de possíveis problemas presentes e passados, para que o serviço seja direcionado ao atendimento de suas necessidades em sua globalidade, e que o atendimento contemple sua individualidade, com suas crenças e valores específicos. Para isso, a empresa interessada na pesquisa pretende conhecer alguns interesses:

Perfil social

A situação familiar de um indivíduo pode interferir em algumas maneiras de convívio ou generalizar seus hábitos com pessoas correspondentes de relacionamento, cultura, comportamento, interesses, etc.

A análise do conteúdo da amostra foi realizada agrupando-se os dados por algumas temáticas: dados acadêmicos, dados pessoais, referências sobre TV aberta local, identificações diretas sobre a programação da Band Curitiba. Foram respondidos 300 questionários no total, conforme o Gráfico 2:

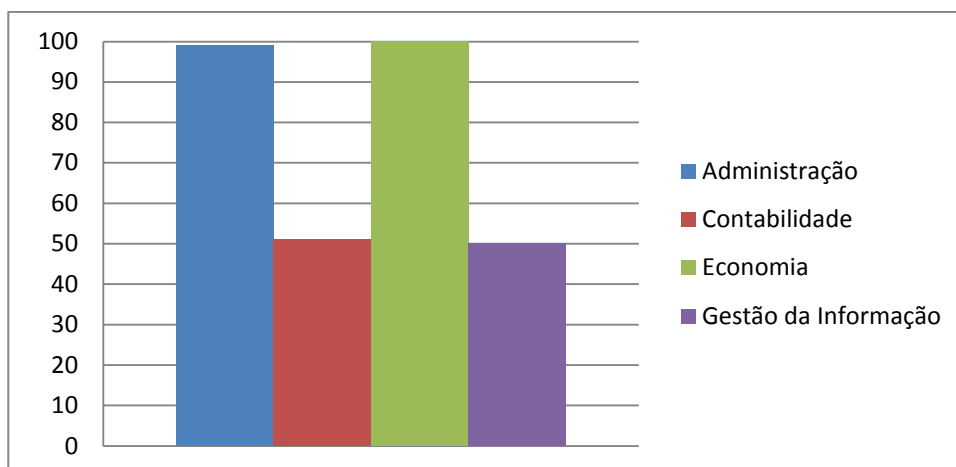


GRÁFICO 2 AMOSTRA DE UNIVERSITÁRIOS DO SA-UFPR
 FONTE: AUTORA COM BASE NOS DADOS DA PESQUISA

De acordo com a amostra, a porcentagem na população do SA-UFPR refere-se a 14,28%.

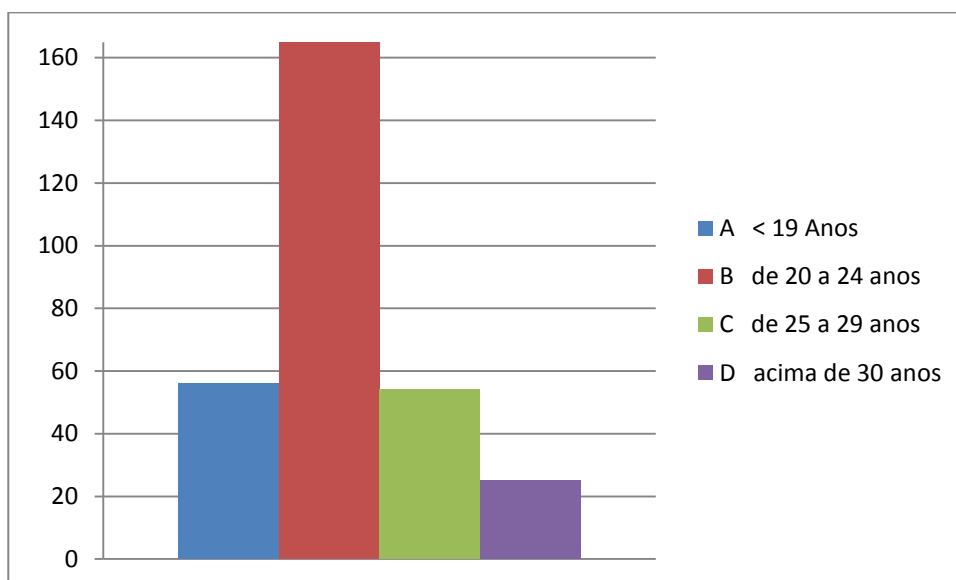


GRÁFICO 3 FAIXA ETÁRIA
 FONTE: AUTORA COM BASE NOS DADOS DA PESQUISA

Pelo Gráfico 2, percebe-se que 55% da amostra possui idade entre 20 e 24 anos. Com relação ao estado civil, 87,3% são solteiros e apenas 10,3% são casados. Relativo ao grau ou período de graduação, 31,6% não relacionou seu tempo de graduação por estarem desperiodizados ou cursando disciplinas isoladas e 38,3% somam-se os de último e penúltimo ano de faculdade. Do total da amostra

pesquisada (N = 300), 52,6% são do sexo masculino e 47,3% do sexo feminino. Este equilíbrio apresenta certa importância quanto ao controle da pesquisa por gênero.

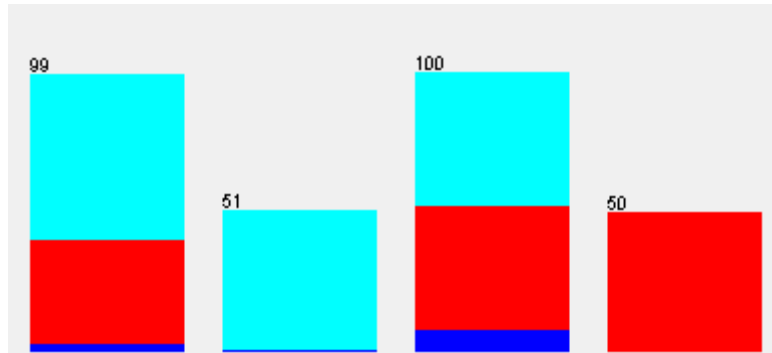


GRÁFICO 4 DISTRIBUIÇÃO DA AMOSTRA POR CURSO E TURNO
FONTE: AUTORA COM BASE NOS DADOS DA PESQUISA

Uma das questões se referia à informação sobre o turno de atividade acadêmica do universitário. O Gráfico 4 apresenta os cursos da esquerda para a direita: Administração, Contabilidade, Economia e Gestão da Informação. As colunas com duas cores referem aos cursos de administração e economia em dois turnos. Colunas com baixo índice na terceira cor referem a turno não informado na entrevista.

Os turnos apontam para um possível horário que torna possível a disponibilização ou tempo para TV, surgindo a oportunidade de investimento nesses horários pré-definidos considerando a faixa etária B de maior incidência conforme os Gráficos 5 e 6. Um fator referente a situação de trabalho do estudante não foi considerado no questionário, o que torna incerto qualquer aplicação com referência na amostra. A distribuição dos cursos por gênero é um importante equilíbrio na amostra apresentada no Gráfico 5. A primeira coluna refere a turno não informado.

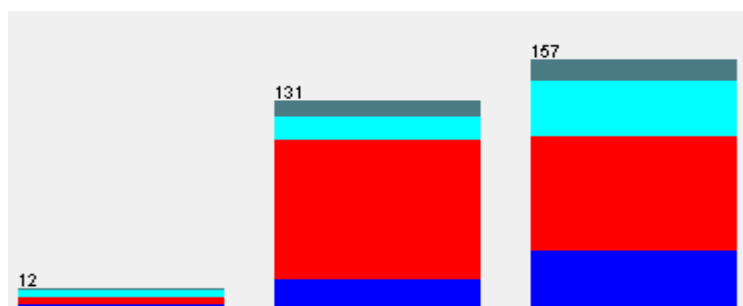


GRÁFICO 5 DISTRIBUIÇÃO DE TURNO POR IDADE
FONTE: AUTORA COM BASE NOS DADOS DA PESQUISA

Com relação à distribuição da faixa etária da amostra pesquisada pode ser visualizada no Gráfico 6. Cada coluna refere a um curso. A faixa etária B (vermelho) predomina para todos os cursos. A indicação das cores pode ser visualizada na Tabela 4. As cores ficam distribuídas assim:

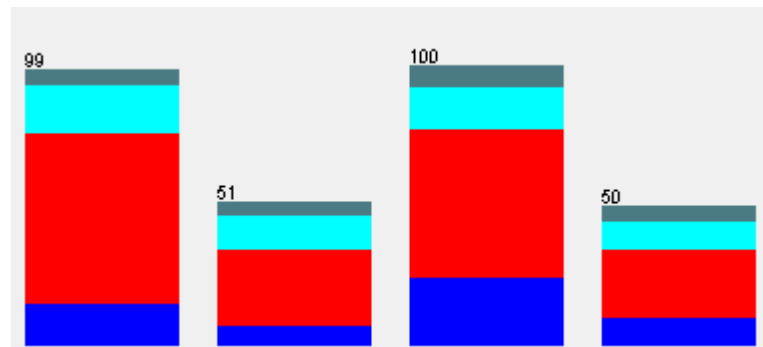


GRÁFICO 6 DISTRIBUIÇÃO POR FAIXA ETÁRIA RELATIVA AOS CURSOS
FONTE: AUTORA COM BASE NOS DADOS DA PESQUISA

TABELA 4: FAIXA ETÁRIA - CURSOS DE GRADUAÇÃO

Faixas etárias		Distribuição por cores
A	(<19 anos)	azul
B	(de 20 a 24 anos)	vermelho
C	(de 25 a 29 anos)	verde-claro
D	(acima de 30 anos)	cinza

FONTE: AUTORA COM BASE NOS DADOS DA PESQUISA

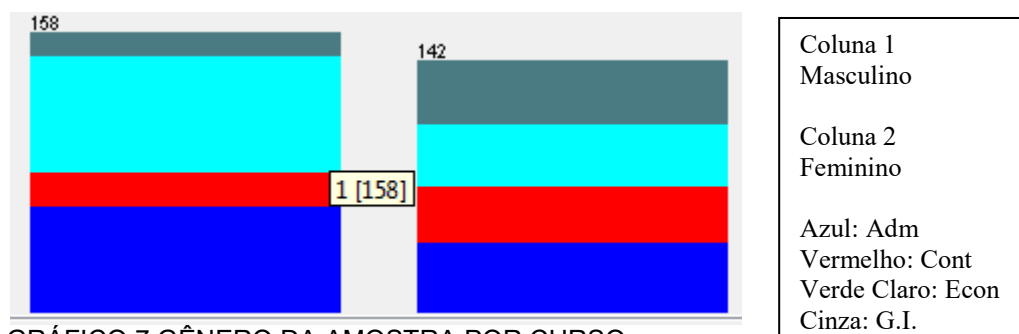


GRÁFICO 7 GÊNERO DA AMOSTRA POR CURSO
FONTE: AUTORA COM BASE NOS DADOS DA PESQUISA

Coluna 1
Masculino

Coluna 2
Feminino

Azul: Adm
Vermelho: Cont
Verde Claro: Econ
Cinza: G.I.

Para atender um público diferenciado por opção individualizada (como para formação acadêmica do indivíduo ou áreas de interesse), este se torna um indicador interessante para hábitos ou elementos distintos para carreira.

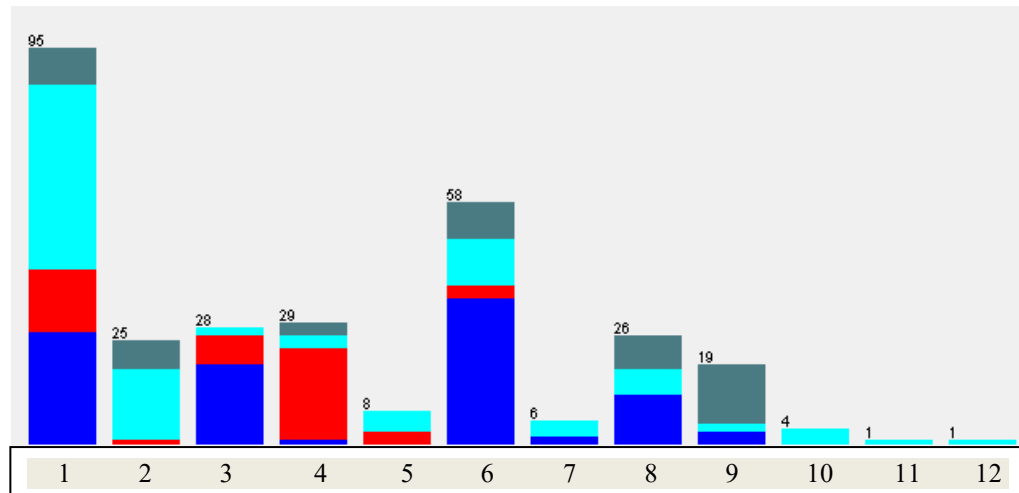


GRÁFICO 8 DISTRIBUIÇÃO POR CURSO SEGUNDO O GRAU ACADÊMICO
 FONTE: AUTORA COM BASE NOS DADOS DA PESQUISA

Cada coluna representa um período de graduação. Os cursos são divididos nas cores: azul (Administração), vermelho (Contabilidade), verde-claro (Economia) e cinza (Gestão da Informação). Os períodos foram classificados da esquerda (com 95 instancias) para direita de acordo com a Tabela 3:

TABELA 3: ORDEM DE PERÍODOS RELATIVOS

1	período sem informação
2	primeiro período
3	segundo período
4	terceiro período
5	quarto período
6	quinto período
7	sexto período
8	sétimo período
9	oitavo período
10	nono período
11	décimo período
12	para 11º período.

FONTE: AUTORA COM BASE NOS DADOS DA PESQUISA

Não existe curso no setor de Ciências Sociais Aplicadas acima do 8º período de graduação. As informações foram relativas ao motivo de desperiodização do estudante. A mesma situação se atribui coluna 1 (período sem informação). Nesta, há possibilidade de os estudantes no último ou penúltimo ano de curso, mas não forneceram tal informação tornando esta classificação duvidosa.

Os indicadores mais importantes são para 5º, 7º e 8º períodos pelo nível de formação acadêmica. Para uma produção de TV voltada para público com esse grau, os entrevistados de maior frequência concentram-se nos cursos de Administração e Gestão da Informação. Quanto a disponibilidade indicada para TV os níveis de audiência são mostrados nos Gráficos 9 e 10 a seguir:

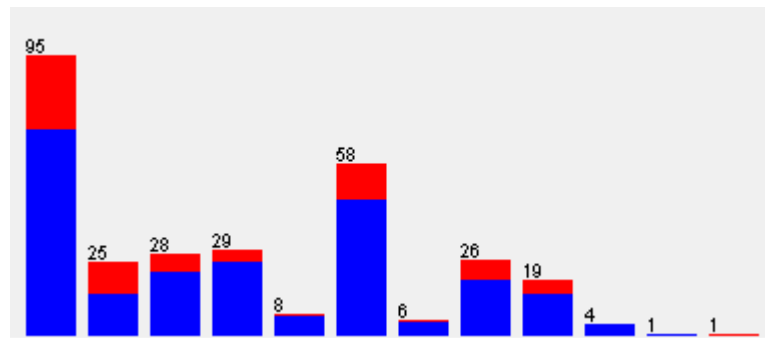


GRÁFICO 9 HORÁRIO PARA TV

FONTE: 1 AUTORA COM BASE NOS DADOS DA PESQUISA

Índice de audiência neste período de graduação concentram-se nos horários das 6h às 7h, das 12h às 13h, 18h às 24h. Os horários podem referenciar uma importante aplicação de negócio com foco em perfil profissional em administração e gestão da informação, por exemplo.

Lazer é um conceito muito atribuído ao tempo livre de um indivíduo. A disponibilidade pessoal, privativa ou coletiva pode indicar uma tendência do indivíduo em tempo aberto para utilização de serviços de TV, por exemplo. Em porcentual, o nível de audiência pode ser visualizado no Gráfico 10:

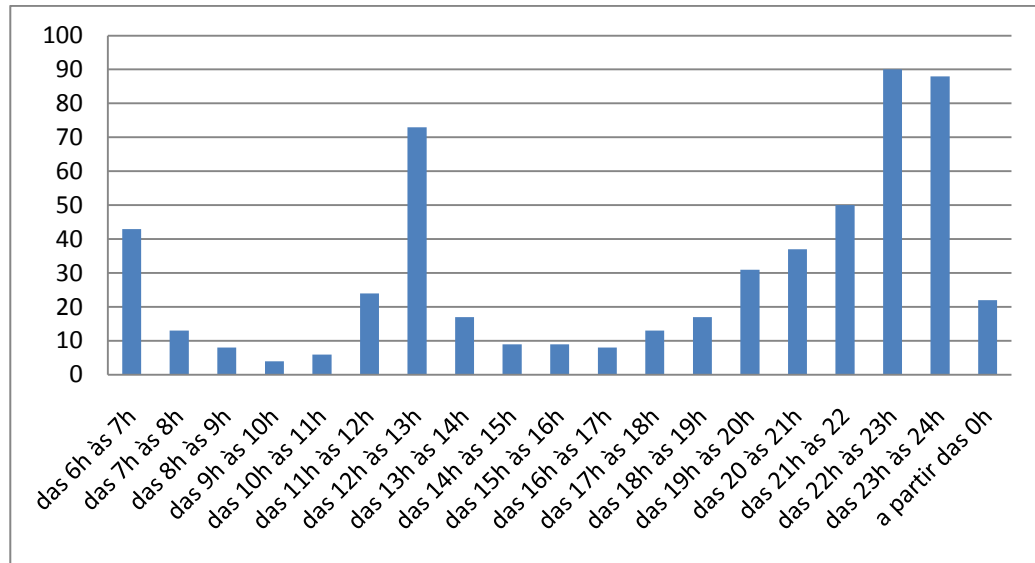


GRÁFICO 10 ÍNDICE DE AUDIÊNCIA

FONTE: AUTORA COM BASE NOS DADOS DA PESQUISA

Para a pesquisa quantitativa, a amostra foi classificada para o perfil sócio econômico do universitário com resultados duvidosos. Em alguns casos, os entrevistados apresentaram dúvida quanto à faixa salarial questionada se renda total familiar ou pessoal. Sendo, esta o total da renda desfrutada ou independência econômica. O resultado não apresenta a real classe social dos entrevistados. Os dados são representados nos gráficos 11 e 12 a seguir:

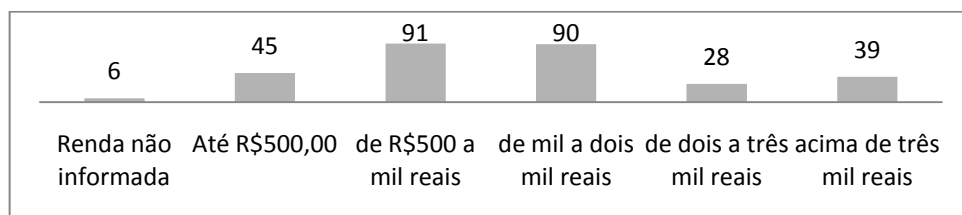


GRÁFICO 11 RENDA

FONTE: AUTORA COM BASE NOS DADOS DA PESQUISA

As divisões A, B, C e D servem para pesquisa de mercado e opinião. A partir do CCEB mantém o objetivo de identificar o que a classe consome para avaliação de expectativas de consumo. A maior incidência de renda é de R\$500,00 (quinhentos reais) a R\$1.000,00 (um mil reais) para o numero de 91 entrevistados e de um a dois mil reais para 90 dos entrevistados. Para o CCEB, o fator não serve

para classificar a amostra em “classes sociais”, mas para indicar possível potencial de compra de alguns consumidores.

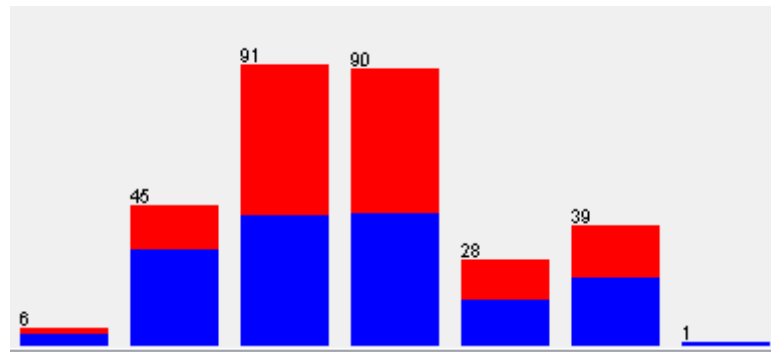


GRÁFICO 12 RENDA POR GÊNERO

FONTE: AUTORA COM BASE NOS DADOS DA PESQUISA

Em porcentual, os apresentadores da TV Aberta Local podem ser visualizados no Gráfico 14:

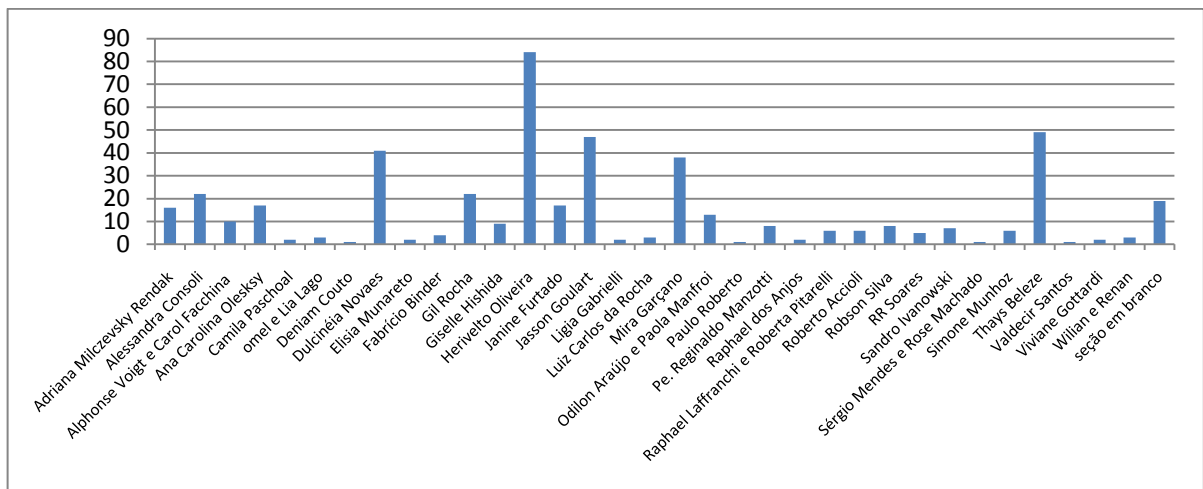


GRÁFICO 13 PREFERÊNCIA POR APRESENTADORES

FONTE: AUTORA COM BASE NOS DADOS DA PESQUISA

Atualmente alguns apresentadores não estão em atividade âncora na TV Band Curitiba como Alphonse e Carol (Jogo Aberto-PR) e Giselle Hishida. Essas foram referências iniciais coletadas desde aplicação da pesquisa no ano de 2009, porém continuaram sendo selecionados pelo possível conhecimento dos entrevistados.

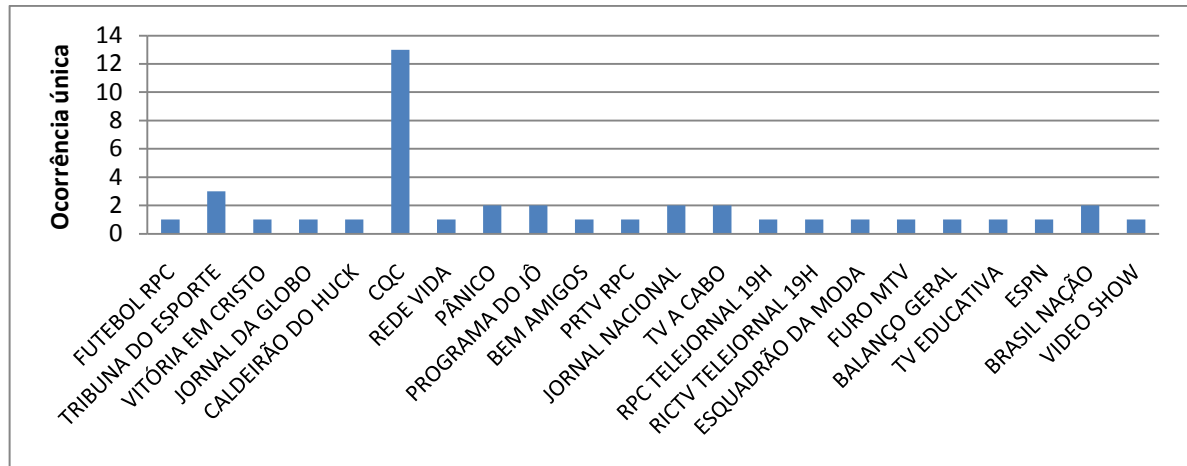


GRÁFICO 14 REFERÊNCIAS A OUTROS PROGRAMAS

FONTE: AUTORA COM BASE NOS DADOS DA PESQUISA

Da mesma forma tratada para apresentadores, a programação nacional foi lembrada por alguns entrevistados como mostra o Gráfico 15. O CQC, programa de televisão de humor apresentado na Rede Bandeirantes, indicou maior audiência em relação aos demais. Depois vem Tribuna do Esporte_programa de esporte da Rede Massa/SBT, Jornal Nacional (tele jornal da Globo) e TV a cabo (um sistema de distribuição de conteúdos audiovisuais de televisão e outros serviços por assinatura).

5.2 RESULTADOS OBTIDOS PELA MINERAÇÃO DE DADOS – FERRAMENTA WEKA

Uma das tarefas escolhidas foi a de descoberta de associações utilizando o algoritmo Apriori disponível na ferramenta WEKA.

O equipamento utilizado foi um computador pessoal: Processador Intel Core Duo com 2GB RAM, executando o sistema operacional Windows Vista Basic.

Após todo o pré-processamento, limpeza e transformação dos dados, a primeira seleção foi o lançamento de dados com 101 atributos candidatos a análise em arquivo formato ARFF para processamento no ambiente WEKA.

Na aplicação do Apriori, usando os 74 atributos em 300 instâncias, a Tabela 5 apresenta os parâmetros utilizados:

TABELA 5: SUPORTE E CONFIANÇA

Suporte Mínimo	Confiança Mínima	Nº de Regras Geradas
0.1 - 10% (30 instancias)	0.9 - 90%	1000

FONTE: AUTORA COM BASE NOS DADOS DA PESQUISA

Apesar da confiança gerada nas 1.000 (um mil) regras, foram selecionadas apenas 531 com confiança de 100% devido a pouca relevância ou resultados óbvios que a técnica Apriori apresentou, como:

Idade=B Convivio4=4 125 ==> Estado_Civil=2 125

Ou

**Sexo=2
Band_Cidade8=8 , PR_Acontece8=8 79 ==> Boa_Tarde_PR5=5 79**

Com esses resultados, foi possível perceber que o número de regras geradas cresce à medida que o suporte diminui, tornando inviável a análise humana mesmo que apresente 100% de confiança. Para facilitar a compreensão, optou-se por trabalhar com conjuntos reduzidos com a seleção de regras consideradas interessantes para análise de perfil do universitário e negócios focados neste público.

5.3 ANÁLISE DAS REGRAS GERADAS

A aplicação do Apriori apresenta às associações em dois subconjuntos distintos reduzidos dos atributos pré-candidatos a análise na quantidade 41 regras podem ser visto a seguir.

Consideraram-se atributos como:

“IDADE”, “SEXO”, “TURNIO”, “HORÁRIOS PARA TV”, “APRESENTADOR(A) PREFERIDO(A)”. A relação pode ser apresentada na Tabela 6:

TABELA 6: MODELO DE REPRESENTAÇÃO DO CONHECIMENTO

L	VARIÁVEIS	Nº de ocorrências	conector	VARIÁVEL CORRESPONDENTE	Nº de ocorrências
1	Idade=B Convivio4=4	125	==>	Estado_Civil=2	125
2	Turno=1 Idade=B Convivio4=4	69	==>	Estado_Civil=2	69
3	Convivio4=4 Apresentador17=17	63	==>	Estado_Civil=2	63
4	Sexo=2 Idade=B Convivio4=4	60	==>	Estado_Civil=2	60
5	Idade=B Faixa_Salarial=2	56	==>	Estado_Civil=2	56
6	Convivio4=4 Horario7=7 Convivio4=4	55 54	==> ==>	Estado_Civil=2	55
7	Apresentador17=17 Boa_Tarde_PR5=5			Estado_Civil=2	54
8	Convivio4=4 Horario19=19 Boa_Tarde_PR5=5	52	==>	Estado_Civil=2	52
9	Sexo=2 Faixa_Salarial=2 Convivio4=4	49 46	==> ==>	Estado_Civil=2	49
10	Faixa_Salarial=3 PR_Acontece8=8			Boa_Tarde_PR5=5	46
11	Faixa_Salarial=1	45	==>	Estado_Civil=2	45
12	Periodo=5 Convivio4=4	45	==>	Estado_Civil=2	45
13	Idade=B Band_Cidade4=4	45	==>	Estado_Civil=2	45
14	Curso=4 Estado_Civil=2	42	==>	Turno=1	42
15	Idade=B Horario7=7	41	==>	Estado_Civil=2	41
16	Turno=2 Convivio4=4 Horario19=19	41	==>	Estado_Civil=2	41
17	Convivio4=4 Horario18=18 jogo_Aberto_PR5=5	41	==>	Boa_Tarde_PR5=5	41
18	Convivio4=4 Horario7=7 PR_Acontece8=8 Boa_Tarde_PR5=5	40	==>	Estado_Civil=2	40
19	Sexo=2 Convivio4=4 Faixa_Salarial=2	38	==>	Estado_Civil=2	38
20	Estado_Civil=2 Horario18=18 Band_Cidade8=8	38	==>	Boa_Tarde_PR5=5	38
21	Estado_Civil=2 Apresentador17=17 Band_Cidade8=8 PR_Acontece8=8	38	==>	Boa_Tarde_PR5=5	38
22	Curso=2 Convivio4=4	37	==>	Estado_Civil=2	37
23	Turno=2 Convivio4=4 Faixa_Salarial=2	37	==>	Estado_Civil=2	37
24	Idade=B Convivio4=4 Band_Cidade4=4	37	==>	Estado_Civil=2	37
25	Convivio4=4 Horario19=19 jogo_Aberto_PR5=5	37	==>	Estado_Civil=2	37
26	Turno=2 Convivio4=4 Horario19=19 Boa_Tarde_PR5=5	37	==>	Estado_Civil=2	37
27	Convivio4=4	36	==>	Estado_Civil=2	36

	Apresentador19=19				
28	Turno=2 Convivio4=4	36	==>	Estado_Civil=2	36
	Band_Cidade4=4				
29	Sexo=2 Convivio4=4	36	==>	Estado_Civil=2	36
	Apresentador17=17				
30	Idade=B Horario7=7	36	==>	Estado_Civil=2	36
	Boa_Tarde_PR5=5				
31	Sexo=2 Idade=B	35	==>	Estado_Civil=2	35
	Faixa_Salarial=2				
32	Curso=4 Convivio4=4	34	==>	Estado_Civil=2	34
33	Convivio4=4 Horario21=21	34	==>	Estado_Civil=2	34
34	Convivio4=4	34	==>	Estado_Civil=2	34
	Apresentador34=34				
35	Convivio4=4	34	==>	Estado_Civil=2	34
	jogo_Aberto_PR1=1				
36	Idade=B	33	==>	Estado_Civil=2	33
	Apresentador19=19				
37	Convivio4=4 Horario17=17	33	==>	Estado_Civil=2	33
38	Idade=B Convivio4=4	33	==>	Estado_Civil=2	33
	Horario18=18				
39	Convivio4=4	30	==>	Estado_Civil=2	30
	Apresentador10=10				
40	Apresentador10=10	30	==>	Boa_Tarde_PR5=5	30
	PR_Acontece8=8				

FONTE: AUTORA COM BASE NOS DADOS DA PESQUISA E FERRAMENTA WEKA

Os resultados apresentados geraram diversas ocorrências com o mesmo apontamento de regras nos atributos. Isso aumenta o grau de certeza em relação aos padrões assimilados. Reduziu-se o número para evitar repetições. A expressão das regras tem a forma $X \Rightarrow Y$ (Se X então Y , ou, se variável então correspondente), ou seja, Y é uma parte condicional da regra (BERRY e LINOFF, 1997). Em análise generalizada, a relação entre o número de transações entre todos os itens e o número de transações somente com itens que fazem parte do „se“, ou seja, é possível relacionar as transações antecedentes por apresentarem mesmo valor de suporte considerável no conjunto de dados. Com essa informação, tem-se a seguinte conclusão de algumas regras relacionadas:

- Segundo a linha 12, estudantes do 5º período de graduação (suporte de 45% da amostra), sexo feminino ou gênero não relacionado (linhas: 4, 9, 19 e 31), idade entre 20 e 24 anos (linhas: 1, 4, 5, 13, 24, 30, 31, 36 e 38), com faixa salarial de R\$500 a R\$1.000,00 (linhas: 5, 11, 23 e 31) estudam pela manhã (linha 2), moram com os pais (linhas: 1, 2, 4,

6, 7, 8, 10, 12, 16, 17, 19, 22, 23, 24, 25, 26, 27, 28, 29, 32, 33, 34, 35, 37, 38 e 39) e são solteiras (indicador correspondente). Essa mesma categoria costuma assistir TV das 12h às 13h (linhas: 15, 18 e 30) e tem preferência pelos apresentadores Herivelton de Oliveira da RPC (linhas: 3, 7, 21, 29 e 37), Jasson Goulart – Globo Esporte RPC (linhas 27 e 36) e Dulcinéia Novaes – Meu Paraná RPC e nunca assistiram, o Boa Tarde PR, obviamente, neste horário ou qualquer programação da Band Curitiba. Esses horários são de considerável incidência para TV (conforme os Gráficos 10 e 14). Se ocorrer aplicação com assuntos similares pela TV Band Curitiba, a divulgação deverá acompanhar o negócio para facilitar a retenção desse público. As mesmas indicações podem ser visualizadas na Figura 8 a seguir:

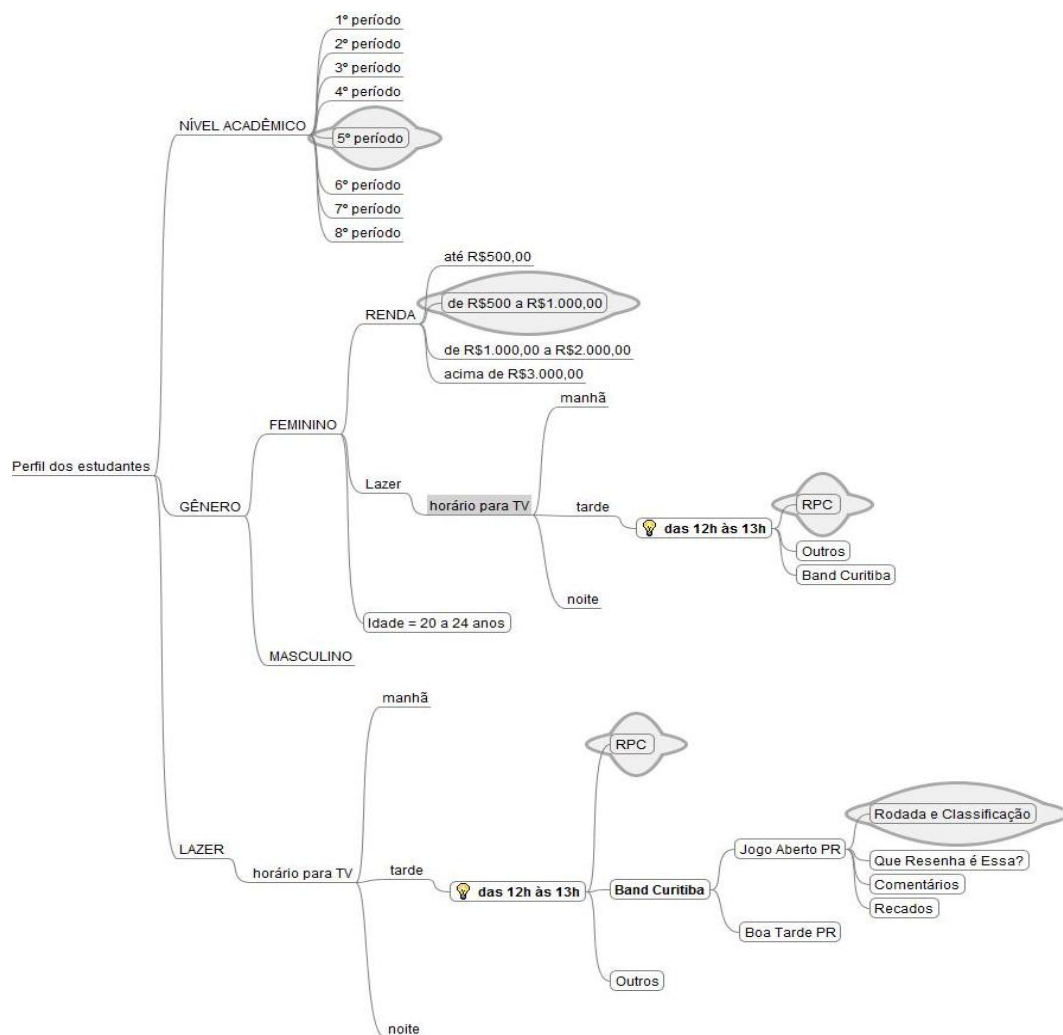


FIGURA 8 MODELO DE REPRESENTAÇÃO DO CONHECIMENTO.
FONTE: AUTORA COM BASE NOS DADOS DA PESQUISA

As visões gerais da informação contida no conjunto de dados indicam nível de graduação, gênero, faixa etária, faixa salarial, turno de atividade acadêmica, convívio familiar, estado civil, horário que costuma assistir TV e apresentadores de preferência e referências sobre a programação da Band Curitiba. O perfil foi descrito, um público perfeitamente definido. Os apresentadores referenciam a competitividade ou como se comporta a concorrência na opinião do público. Estudantes que não assistem a TV Band neste horário podem não ter conhecimento sobre a programação ou o conteúdo de audiência pode indicar um bom investimento nos horários atribuídos.

- Frequência de padrões com atributos da programação local sem referência ou nenhum conhecimento pelos entrevistados, como:

**Sexo=2 Estado_Civil=2 Band_Cidade8=8 jogo_Aberto_PR5=5 65 ==> Boa_Tarde_PR5=5 65
conf:(1)**

Dentre mil regras geradas, mais de 60% afirmaram nunca terem assistido qualquer programação oferecida pela Band Curitiba. Isso pode ocorrer devido ao serviço diferenciado apresentado pela TV Band Curitiba ou porque os estudantes universitários gostam de notícias e esporte, apenas. As regras não foram apresentadas nesta seção para evitar repetições demonstração em excesso.

- Estudantes solteiros ganham até de R\$1.000,00 (linhas: 5, 23, e 31), estudam a noite (linhas: 16, 23, 26 e 28) e são solteiros e moram com os pais (linhas: 1, 2, 4, 6, 7, 8, 10, 12, 16, 17, 19, 22, 23, 24, 25, 26, 27, 28, 29, 32, 33, 34, 35, 37, 38 e 39). Uma das regras geradas com indicador objetivo para esta informação teve um suporte de 33% da amostra. A transação pode ser visualizada na linha 23 e melhor representada na Figura 9:

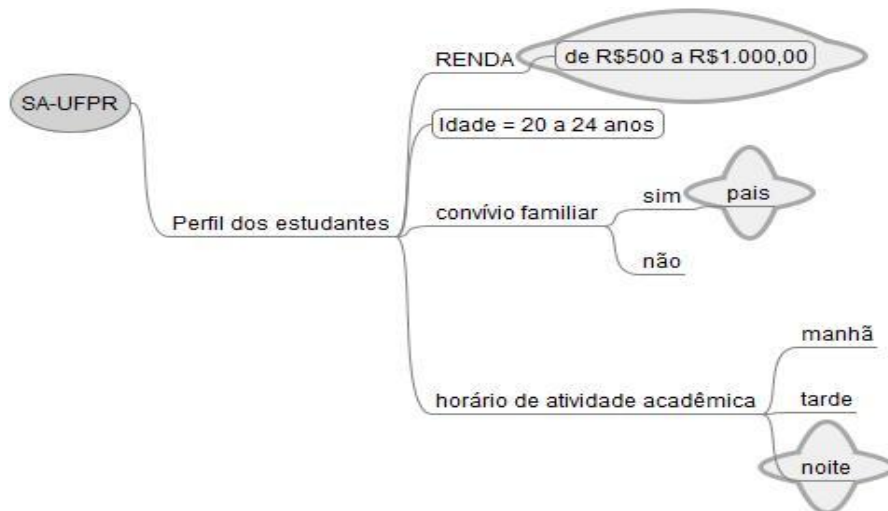


FIGURA 9 MODELO DE REPRESENTAÇÃO DO CONHECIMENTO
 FONTE: AUTORA COM BASE NOS DADOS DA PESQUISA

A faixa salarial e faixa etária foram importantes critérios solicitados pela empresa interessada. O indicador refere a uma classe social permitindo maior definição de exigência ou consumo de conteúdo.

- Estudantes solteiros com idade de 20-24 (linhas: 24, 28 e 39), que estudam a noite, preferem assistir o quadro de esporte oferecido pelo Band Cidade ou outro canal com mesmo assunto, conforme a Figura 10:

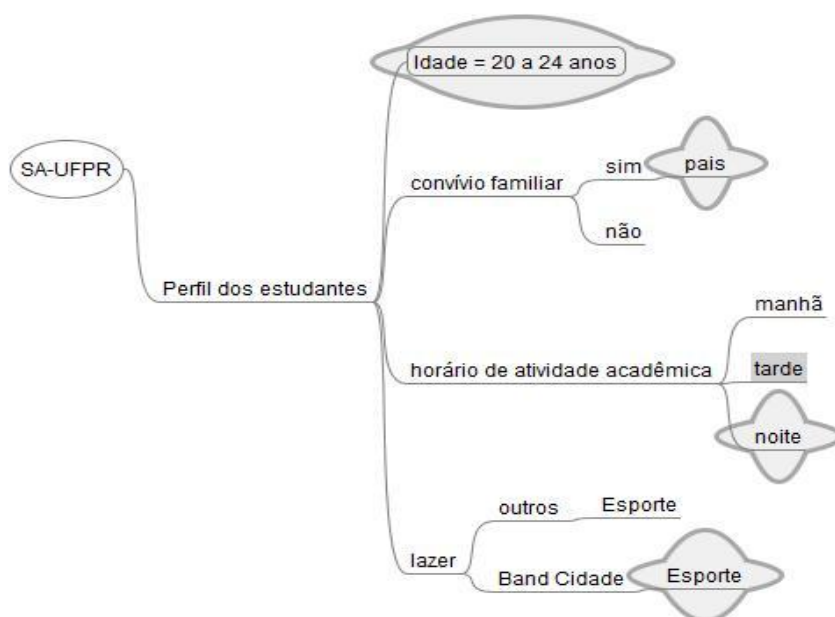


FIGURA 10 REPRESENTAÇÃO DO CONHECIMENTO
 FONTE: AUTORA COM BASE NOS DADOS DA PESQUISA

Assunto relacionado ao esporte foi credibilidade reconhecida para TV Band Curitiba. A competitividade pode ser identificada aqui como referência regional. Para atingir esse público, quadros ou programas sobre Esporte devem ter produção mais voltada para o perfil masculino com incidência em horários das 12h às 14h. Outros padrões podem associar a mesma frequência:

- Estudantes solteiros que moram com os pais preferem o quadro de Rodadas e Classificação oferecido pelo Jogo Aberto PR (linha 35) com suporte de 34%;
- Estudantes de idade entre 20 e 24 anos, solteiros preferem a apresentação de Jasson Goulart – Globo Esporte RPC (linhas 27 e 36).
- Estudantes solteiros e que moram com os pais assistem TV a partir das 22h, maior incidência nos horários das 23 à 24h (linhas: 8, 17, 20 e 25) numa frequência de 52% a 37% da amostra.
- Outra apresentadora de maior preferência é a Dulcinéia Novaes (Meu Paraná – RPC), (linhas: 39 e 40). Para superar a audiência no horário desse programa concorrente, a programação da TV Band Curitiba deve ter produção similar.

A grande relação entre quem assiste TV e tem preferência pelos apresentadores de outra transmissora pode evidenciar um público já consolidado ou qualidades referentes à apresentação da TV concorrente. O espaço para essa opinião permitiu a avaliação da audiência concorrente em horários definidos. Para que a TV Band atue com superação no mercado, é preciso que adicione em sua produção assuntos similares ou inove sua programação com conteúdo competitivo para o público jovem universitário, por exemplo. Em geral, este é um público que prefere notícias.

6 CONSIDERAÇÕES FINAIS

A pesquisa proporcionou conceituar e aprofundar estudos com a compreensão de todos os passos do KDD como uma atividade profissional na identificação de inconsistências, com considerável atenção a valores ausentes que interferem na imputação de dados ou resultados de importância. Em seguida a abordagem de técnica de Mineração de Dados para escolha de melhor tarefa para obtenção de conhecimento a partir de problemas propostos em estudo de caso. O método Apriori foi utilizado para tratar de problemas reais e semelhantes com resultados diferentes devido ao método utilizado.

Com o objetivo geral atendido, foi completa a realização de um sistema convincente. Perfis foram comportados em conjunto (ou base de dados) indicando clientes telespectadores para descoberta de conhecimento através de associações interessantes a fim de apresentar informação útil à tomada de decisão e com auxílio da aplicação do algoritmo Apriori da Mineração de Dados. Especificamente foram identificados tipos de telespectadores e apresentados seus perfis de acordo com suas preferências para TV; foi avaliado se a programação de TV de estudo de caso é adequada às preferências e disponibilidade dos telespectadores; verificou-se o acesso de telespectadores a canais de TV diferenciados e/ou para assuntos específicos; foi percebido o nível de frequência diária dos telespectadores na TV através das regras de Mineração de Dados para obter informação de valor futuro ou desconhecido. As regras de associação não permitem indicar valor futuro ou predição pra alguma ocorrência, mas as associações sugerem uma ação, uma possível probabilidade.

A realização do estudo de caso foi possível graças a esse procedimento que favorece o estudo sobre clientes, sobre ambiente e também a identificação de novas oportunidades proporcionadas pela da pesquisa. A dificuldade encontrada neste trabalho foi mais a aplicação da pesquisa do que a identificação ou análise dos tipos de dados. A entrevista foi decorrente de vários fatores que podem favorecer ou prejudicar a pesquisa contando com disponibilidade, aceitação, fatores emocionais e cultura em relação à colaboração no estudo.

Mas o conhecimento desejado intrínseco na relação dos telespectadores universitários do SA-UFPR favoreceu a identificação e análise de tipos de dados até antes desconsiderados e visualizados através da classificação da informação,

limpeza dos dados e validação dos resultados. Em essência, é para o reconhecimento da utilização de recursos de informação que as empresas devem perceber essa importância como contribuição para um fator crítico de sucesso. Para a IC, a correta utilização é que torna uma empresa inteligente para negócios, ou seja, desenvolver solução que possibilite a coleta de informações no ambiente externo à organização, uma forma de monitorar o ambiente. Assim, a Mineração de Dados contribui para esse auxílio ao planejamento da organização, aprendizado e tomada de decisão com informação encontrada em seu processo de Descoberta de Conhecimento para vantagem competitiva.

Nesta pesquisa, fizeram-se algumas conexões com as fronteiras de desenvolvimento de tecnologias informacionais que são relevantes à análise de negócio para o serviço de mídia na TV, a BI. Nesta, a proposta é para que contribua para uma possível expansão de infra-estrutura de sistemas de informação em armazenamento ou relacionamento de dados com utilização da técnica de Mineração de Dados. A sociedade de informação obriga as organizações a acompanhar algumas transformações tornando necessário a incorporação de algumas tecnologias obrigando uma relação mais estreita com as ciências da computação desenvolvendo produtos ou oferecendo serviços com auxílio da IA para usar as tecnologias informacionais em análises como o perfil do telespectador e sua satisfação, ou seja, aplicação de conceitos e metodologias de BI.

Transformar dados externos ou coletados em informação para organização, definir ou apresentar resultados como indicadores de negócios, reduzir custos com aquisições de serviços externos, contribuir para o desempenho ou projetar crescimento alinhando planos e metas, proporcionais ganhos de produtividade têm sido objetivos extremamente necessários para o mundo empresarial e competitivo.

Não basta as empresas possuírem grandes bases de dados, ou extrair só a informação essencial do mercado, como nesta pesquisa, a amostra é apenas uma fatia de estudantes num universo acadêmico como é a cidade de Curitiba. O objetivo da *Business Intelligence* é antecipar mudanças de mercado e concorrente com inteligência, ou seja, favorecer o aprendizado, rever suas próprias práticas organizacionais ou de negócios e inovar.

Sem modismos ou fazer com que se acredite que as tecnologias de informação possam fazer mais do que elas fazem na prática, a realidade é que existem diversos serviços com alta precisão como aparelhos que funcionam como

transmissor em tempo real de uma dada situação, o desafio é tornar todos os dados recebidos de modo a melhorar o acesso e descobrir algo desconhecido ou precioso como o próprio nome *mineração* pode atribuir a tal idéia. O interessante é pensar que no futuro, essa inteligência artificial possa contribuir não só para geração de lucros e sucesso nos negócios, mas sim contribuir para o desenvolvimento de nossa ciência, quem sabe, através da satisfação de um simples consumidor.

A solução apresentada possui algumas contribuições importantes que é o descobrimento do tipo de cliente como público universitário que a TV Band Curitiba desconhecia. Esta pesquisa foi possível através de métodos convencionais, com estudo e pouco custo envolvido. Através dos padrões Apriori os valores puderam ser descobertos formando uma base consistente de informações descritas em ocorrências reais e dados novos. Com isso possibilitou ao usuário, analista de dados e, principalmente, ao departamento de marketing ter as informações atualizadas e maior controle sobre seu público que já apresentavam distantes dessa oferta que, ultimamente, vem investindo em qualidade e serviço com tecnologia para satisfação do usuário.

6.1 ALGUMAS CONTRIBUIÇÕES

As propostas de trabalhos futuros que podemos citar são:

- a) Comparar os resultados obtidos no Weka com uma ferramenta comercial;
- b) Realizar aplicações de Mineração de Dados com outros algoritmos.
- c) Validar o resultado obtido com especialista da área de programação de TV pesquisada;
- d) Contribuir para aplicações *Business Intelligence* e Inteligência Competitiva a partir de conjuntos de dados integrados em organizações;
- e) Permitir análise das mídias de TV.

REFERÊNCIAS

ABEP. Associação Brasileira de Empresas de Pesquisas: critério de classificação econômica do Brasil. Disponível em: <www.abep.org>

AGRAWAL, R; IMIELINSKI, T; SWAMI, A. *Mining Association Rules Bet-ween Sets of Itens in Large Databases*. ACM SIGMOD Conference Management of Data, 1993.

ALMEIDA, Adiel; COSTA, Ana P. Sistemas de Informação e Gestão do conhecimento. Introdução à Engenharia de Produção. Ed.2008. **Journal Sistemas & Gestão**, v.3, n.3, 2008.

BARBIERI, Carlos. BI. **Business Intelligence – Modelagem e Tecnologia**. Rio de Janeiro, Axcel Books, 2001.

BARRETO, Auta Rojas. A informação eficaz na empresa. **Ciência da Informação**, Brasília, v. 20, n. 1, p. 78-81, jan./jun. 1991.

Batista, G. E. A. P. A. (2003). **Pré-processamento de Dados em Aprendizado de Máquina Supervisionado**. Tese de Doutorado, Instituto de Ciências Matemáticas e de Computação - USP São Carlos.

BERRY, Michael J. A, LINOFF, Gordon. **Data mining techniques**: for marketing, sales, and customer support. New York : J. Wiley E Sons, 1997.

CARVALHO, Luís A. **Datamining** : a mineração de dados no marketing, medicina, economia, engenharia e administração. São Paulo: Ciencia Moderna, 2005.

CHOO, C. W. **Information management for the intelligent organization**: the art of scanning the environment. 3 ed. Medford: Information Today, 2002.

DAVENPORT, T.H. **Ecologia da Informação**: porque só a tecnologia não basta para o sucesso na era da informação. São Paulo: Futura, 1998a. p. 18.

DAVENPORT, Thomas H.; PRUSAK, Laurence. **Conhecimento empresarial**: como as organizações gerenciam o seu capital intelectual. Rio de Janeiro: Campus, 1998.

GIL, Antônio Carlos. **Como elaborar projetos de pesquisa**. 3 ed. São Paulo: Atlas, 1996. 159p.

GLOBO. TV Globo. Disponível em: < <http://g1.globo.com>>.

GOLDSCHMIDT, R. R. ; PASSOS, E. P. L. . **Data Mining**: um guia prático - conceitos, técnicas, ferramentas, orientações e aplicações. 1. ed. Rio de Janeiro: Editora Campus, 2005. v. 1. 250p.

GOMES, Elisabeth e BRAGA, Fabiane (2004). **Inteligência Competitiva**: Como transformar informação em um negócio lucrativo. Rio de Janeiro : Campus, 2005?

GONÇALVES, Lóren Pinto Ferreira. **Avaliação de Ferramentas de Mineração de Dados como fonte de dados relevantes para a tomada de decisão**. 2001. 104 f. Dissertação (Mestrado em administração) – Escola de Administração, Universidade federal do Rio Grande do Sul, Porto Alegre, 2001.

IBGE. **Censo Demográfico 2010**: Questionário: descrição das variáveis. Disponível em: <http://www.ibge.gov.br>. Acesso em: 18 out. 2009.

IBGE. **Geociências**: Cartografia. Área territorial oficial, pasta 2009. Disponível em: <http://www.ibge.gov.br>. Acesso em: 19 jun. 2010.

IBOPE MÍDIA. **Sustentabilidade**: Metodologia de pesquisas de Audiência de TV. Disponível em: <http://www.ibope.com.br>. Acesso em: 18 out. 2009.

FAYYAD, Usama M.; PIATETSKY-SHAPIO, Gregory; SMYTH, Padhraic. **Advances in Knowledge Discovery and Data Mining**. Menlo Park, EUA: AAAI Press, 1996. 611 p.

KERLINGER, F. N. **Metodologia da pesquisa em ciências sociais**. São Paulo: EPUEDUSP, 1980.

KLÖSGEN, Willi (Org.). **Handbook of Data Mining and Knowledge Discovery**. New York, EUA: Oxford University Press, 2002.

MALHOTRA, Naresh. **Pesquisa de Marketing**: uma orientação aplicada. Rio Grande do Sul: Artmed Editora S.A, 2001.

MANUAL de gestão de serviços de informação. Curitiba: TECPAR/ Brasília: IBICT, 1997. 257 p.

MARCHIORI, Z.P., A ciência e a gestão da informação: compatibilidade no espaço profissional. **Ci. Inf.**, Brasília, v. 31, n. 2, p. 72-79, maio/ago. 2002.

MARCONI, M. D. A.; LAKATOS, E. M. **Técnicas de pesquisa**: planejamento e execução de pesquisas, amostragens e técnicas de pesquisas, elaboração, análise e interpretação de dados. 3.ed. São Paulo: Atlas, 1996.

NAVEGA, Sergio C. **Princípios essenciais do Data Mining**. Inteliwise Publicações, 2002. Disponível em <http://www.inteliwise.com/snavega>

RÁDIO E TELEVISÃO EDUCATIVA DO PARANÁ. Brasil Nação. Disponível em: <<http://www.rtve.pr.gov.br>>.

REDE BANDEIRANTES. **Televisão Bandeirantes do Paraná**: Planejamento e Marketing, Band Curitiba. Curitiba, 2009.

REZENDE, Denis A.; ABREU, Aline França de. **Tecnologia da Informação aplicada a sistemas de informação empresariais**. 3 ed. São Paulo: Atlas, 2003.

RIBEIRO, Jeferson; et al. O impacto na utilização de inteligência de negócios no ambiente empresarial. In: Congresso Brasileiro Gestão e Desenvolvimento de Produtos, IV, 2003, Gramado. **Resumos...** Gramado, 2003.

RICK, Elaine; KNIGHT, Kevin. **Inteligência Artificial**. São Paulo: Makron Books, 1993.

RICK, Elaine; KNIGHT, Kevin. **Inteligência Artificial**. São Paulo: Makron Books, 1993. p.554

SELLTIZ, C. et al. **Métodos de pesquisa nas relações sociais**. 3 ed. São Paulo: E.P.U., 1974.

TAN, P.N; STEINBACK, M.; KUMAR, V. **Introdução ao Data Mining**. Rio de Janeiro: ed. Ciência Moderna, 2009.

TERRA, José Cláudio Cyrineu (2003). **Gestão do Conhecimento e ELearning na prática**. Rio de Janeiro : Elsevier, 2003.

TERRA, José Cláudio Cyrineu e ALMEIDA, Carolina (2008). **Gestão do Conhecimento e Inteligência Competitiva**: duas faces da mesma moeda. Biblioteca Terra Forum Consultores. Disponível em <<http://www.terraforum.com.br/sites/terraforum/Biblioteca/Duas%20faces%20da%20mesma%20moeda.pdf>>. Acesso em: 15 out de 2009.

TV IGUAÇU. Rede Massa: programação. Sistema Brasileiro de Televisão. Disponível em: < <http://www.redemassa.com.br>>. Acesso em: 10 set. 2009.

UOL. **Michaelis Moderno Dicionário da Língua Portuguesa**. São Paulo: Melhoramentos Ltda. 2009. Disponível em <<http://michaelis.uol.com.br>>. Acesso em 18 out. 2009.

UOL NOTÍCIAS. Ooops. Disponível em: < <http://noticias.uol.com.br/ooops/ultnot/2009/12/21/ult2548u820.jhtm>>

UNIVERSIDADE FEDERAL DO PARANÁ. Informações sobre a Universidade. Disponível em: <http://www.ufpr.br/adm/templates/p_index.php?template=3&Cod=79&hierarquia=6.1>. Acesso em 20 de novembro de 2009.

WEKA 3: **Data Mining Software in Java**. Disponível em <<http://www.cs.waikato.ac.nz/ml/weka/index.html>>. Acesso em: 2 dez 2009.

WITTEN, Ian, H., Frank E., **Data Mining – Practical Machine Learning Tools and Techniques with Java Implementations**. Morgan Kaufmann, 2000.

APÊNDICE A FORMULÁRIO – PERFIS DE TELESPECTADORES DE TV ABERTA LOCAL ENTRE UNIVERSITÁRIOS DO SETOR DE CIÊNCIAS SOCIAIS APLICADAS DA UFPR

Data : ____ / ____ / 2009		Formulário N° ____
Curso:	1 - <input type="checkbox"/> Administração	3 - <input type="checkbox"/> Economia
	2 - <input type="checkbox"/> Contabilidade	4 - <input type="checkbox"/> Gestão da Informação

Turno :	1 - <input type="checkbox"/> Matutino	2 - <input type="checkbox"/> Noturno	Período: _____
---------	---------------------------------------	--------------------------------------	----------------

1 CARACTERÍSTICAS GERAIS DO UNIVERSITÁRIO

1.0 – NOME E TELEFONE (*preenchimento opcional*):

(0xx41): _____

1.01 – SEXO:

1 - ☐ masculino 2 - ☐ feminino

1.02 – DATA DE NASCIMENTO:

1 - ____ / ____ / 19 ____

1.06 – ESTADO CIVIL:

1 - ☐ casado (a)

4 - ☐ separado (a)

2 - ☐ solteiro (a)

5 - ☐ união estável

3 - ☐ viúvo (a)

1.06A – PESSOAS DO CONVÍVIO:

1 - ☐ marido/companheiro (a)

4 - ☐ pais

2 - ☐ filhos

5 - ☐ outros _____

3 - ☐ só

1.07 – FAIXA SALARIAL:

1 - ☐ Até R\$500,00

12 - ☐ de R\$2.000,00 a R\$3.000,00

2 - ☐ de R\$500,00 a R\$1.000,00

13 - ☐ acima de R\$3.000,00

3 - ☐ de R\$1.000,00 a R\$2.000,00

1.08 – Natural de: _____

1.09 – RESIDE NA CIDADE DE: _____

2	PREFERÊNCIAS DO UNIVERSITÁRIO EM RELAÇÃO À TV ABERTA LOCAL
----------	---

2.01 – EM QUAL(S) HORÁRIO(S) VOCÊ COSTUMA ASSISTIR A PROGRAMAÇÃO LOCAL?
Manhã

- 1 - ☐ das 6 às 7h
- 2 - ☐ das 7 às 8h
- 3 - ☐ das 8 às 9h
- 4 - ☐ das 9 às 10h
- 5 - ☐ das 10 às 11h
- 6 - ☐ das 11 às 12h

Tarde

- 7 - ☐ das 12 às 13h
- 8 - ☐ das 13 às 14h
- 9 - ☐ das 14 às 15h
- 10 - ☐ das 15 às 16h
- 11 - ☐ das 16 às 17h
- 12 - ☐ das 17 às 18h

Noite

- 13 - ☐ das 18 às 19h
- 14 - ☐ das 19 às 20h
- 15 - ☐ das 20 às 21h
- 17 - ☐ das 21 às 22h
- 18 - ☐ das 22 às 23h
- 19 - ☐ das 23 às 24h
- 20 - ☐ a partir das 0h

2.02 – POR QUAL(S) APRESENTADOR(S) E/OU PROGRAMA(S) VOCÊ TEM PREFERÊNCIA?:

38. Adriana Milczewsky Rendak
(Mapa Tempo - Bom Dia
Paraná – RPC)

39. Alessandra Consoli (Bom Dia
Paraná – RPC TV)

40. Alphonse Voigt e Carol Facchina
(Jogo Aberto Paraná – Band)

41. Ana Carolina Oleksy (Previsão
do Tempo – RPC)

42. Antônio Carlos da Silva (Carros
& Motores – TV Iguaçu)

43. Camila Paschoal (Jornal da
Educativa– TVE PR)

44. Cecília Comel e Lia Lago
(Destaque – TV Iguaçu)

45. Cesar Camargo e por Kelly
(TransOfertas – TV
Transamérica)

46. Denian Couto (Jornal da Massa
– TV Iguaçu)

47. Dulcinéia Novaes (Meu Paraná
– RPC)

48. Elisia Munareto (Enfoque – TVE
PR)

49. Eugênio Stefanello e Lauren
Trilha (Negócios da Terra – TV
Iguaçu)

50. Fabrício Binder (Band Cidade e
PR Acontece)

51. Gil Rocha (Revista RPC)

52. Giselle Hishida (Band Cidade)

53. Giselle Macedo (Nossa Tarde -
Band)

54. Herivelto Oliveira (ParanáTV –
RPC TV)

55. Janine Furtado (Plug – RPC)

56. Jasson Goulart (Globo Esporte
- RPC)

57. Lígia Gabrielli (Jornal Mercosul –
Rede Mercosul)

58. Luiz Carlos da Rocha (Jogo do
Poder – CNT)

59. Mira Graçano (ParanáTV –
RPC TV)

60. Odilon Araújo e Paola Manfro
(Revista RPC)

61. Paulo Roberto (Boa Tarde PR -
Band)

62. Pe. Reginaldo Manzotti
(Palavras de Fé TV Iguaçu e
Band Curitiba)

63. Raphael dos Anjos (Paraná no
Ar - RIC TV)

64. Raphael Laffranchi e Roberta
Pitarelli (RIC Notícias)

65. Roberto Accioli (190 – CNT)

66. Robson Silva (Tribuna da
Massa 2ª Ed – TV Iguaçu)

67. RR Soares (Show da Fé – TV
Iguaçu)

68. Sandro Ivanowski (Caminhos
do Campo – RPC)

69. Sérgio Mendes e Rose
Machado (RIC Rural)

70. Simone Munhoz (RIC Notícias)

71. Thays Beleze (ParanáTV –
RPC TV)

72. Valdecir Santos (Tribuna da
Massa 1ªEd – TV Iguaçu)

73. Viviane Gottardi (Vitrine
Mercosul – Rede Mercosul)

74. Wiliam e Renan (TV Educativa)

75. Outro (os) apresentador (s):

76. Outro (os) programa (s):

3	PREFERÊNCIAS DO UNIVERSITÁRIO EM RELAÇÃO À BAND CURITIBA
----------	---

3.03 – POR QUAL(S) ASSUNTO(S) VOCÊ TEM PREFERÊNCIA DENTRO DA PROGRAMAÇÃO DA BAND CURITIBA E/OU QUAL SUA OPINIÃO EM RELAÇÃO AOS TEMAS APRESENTADOS?

3.03A – BAND CIDADE (noticiário de 2ª a sábado às 18h50min)

1 - <input type="checkbox"/> – Política	4 - <input type="checkbox"/> – Esporte	7 - <input type="checkbox"/> – Segurança
2 - <input type="checkbox"/> – Vida Pública	5 - <input type="checkbox"/> – Saúde	
3 - <input type="checkbox"/> – Cultura	6 - <input type="checkbox"/> – Educação	
8 - <input type="checkbox"/> – Nunca assisti		
9 - <input type="checkbox"/> – <u>opinião e/ou crítica:</u> _____ _____ _____		

3.03B – PR ACONTECE (noticiário de 2ª a sexta às 6h50min)

1 - <input type="checkbox"/> – Política	4 - <input type="checkbox"/> – Esporte	7 - <input type="checkbox"/> – Segurança
2 - <input type="checkbox"/> – Vida Pública	5 - <input type="checkbox"/> – Saúde	
3 - <input type="checkbox"/> – Cultura	6 - <input type="checkbox"/> – Educação	
8 - <input type="checkbox"/> – Nunca assisti		
9 - <input type="checkbox"/> – <u>opinião e/ou crítica:</u> _____ _____ _____		

3.03C – BOA TARDE PR (telejornal de 2ª a 6ª-feira às 13h30min com Paulo Roberto)

1 - <input type="checkbox"/> – <u>Policial</u> : a sessão acompanha e investiga ocorrências envolvendo crimes contra a sociedade como violência, assassínatos e drogas.
2 - <input type="checkbox"/> – <u>Solidariedade</u> : atende pedidos de pessoas em situação financeira precária com problemas de saúde e necessidades especiais buscando apoio com doações, parcerias voluntárias e solução direto aos órgãos públicos.
3 - <input type="checkbox"/> – <u>Lig Band</u> : Portal de voz que permite a comunicação do telespectador com o programa gravando reivindicações, manifestação de apoio, recados, críticas, e etc.
4 - <input type="checkbox"/> – <u>Me Chama Que Eu Vou</u> : através do Lig Band, o telespectador grava seu recado solicitando visita do apresentador no ambiente urbano/bairro com situações precárias de atendimento por parte dos órgãos públicos como em casos de saneamento, saúde, educação, segurança, etc.
5 - <input type="checkbox"/> – Nunca assisti
6 - <input type="checkbox"/> – <u>opinião e/ou crítica:</u> _____ _____ _____

3.04D – JOGO ABERTO PR (programa de esporte de 2ª a 6ª-feira às 12h30min com Alphonse Voigt e Carol Facchina)

1 - <input type="checkbox"/> – <u>Rodada e Classificação</u> : apresenta posições em campo e a classificação dos jogos de futebol em destaque na semana.
2 - <input type="checkbox"/> – <u>Que Resenha é Essa?</u> : Apresenta algumas declarações e imagens inconvenientes em edições especiais com destaque para a crítica com humor.

3 - ☐ – Recados : Participações por SMS, Twiter e Orkut.

4 - ☐ – Comentários: Discussão com participação do telespectador referente aos jogos da semana ou ao o que tem acontecido no esporte (futebol principalmente) com comentaristas como Rodrigo Meister, Hassan Neto, Fernando Lisboa, Leonardo Portes e Caxias.

5 - ☐ – Nunca assisti

6 - ☐ – opinião e/ou crítica: _____

3.05E – NOSSA TARDE (programa familiar de 2ª a 6ª-feira às 13h)

1 - ☐ – Profissões: apresenta atividades profissionais como orientação e/ou incentivo para formação técnica ou alternativa para fonte de renda.

2 - ☐ – Saúde: espaço dedicado à discussão e orientação com profissionais especializados apresentados em estúdio ou acompanhamento no tratamento de doenças em alguns casos.

3 - ☐ – Viajando com a Gi: permite que o telespectador tenha a oportunidade de realizar viagens turísticas com a apresentadora através de uma história que justifique seu interesse e desejo em viajar com a Produção.

4 - ☐ – Jogo da Tarde: espaço para entretenimento com jogos e prêmios.

5 - ☐ – Momento Místico: espaço para entretenimento com taro terapeutas, tarólogos e búzios.

6 - ☐ – Cultura: eventualmente apresenta bandas, cantores solo e diversos artistas para apresentações.

7 - ☐ – Promoções: realiza sorteios para eventos, jantares e outros prêmios patrocinados.

8 - ☐ – Nunca assisti

9 - ☐ – opinião e/ou crítica: _____

3.06F – BAND PÉ NO RIO (programa universitário apresentado aos sábados às 19h15min)

1 - ☐ – Competição entre equipes de universitários em cenários com rios, mananciais e projetos sociais nos bairros de Curitiba e Região Metropolitana com a missão de informar, entreter e promover ações de preservação do meio ambiente.

8 - ☐ – Nunca assisti

9 - ☐ – opinião e/ou crítica: _____
